Hartmut Kliemt*

Economic and Sociological Accounts of Social Norms

https://doi.org/10.1515/auk-2020-0003

Abstract: Classifying accounts of institutionalized social norms that rely on individual rule-following as 'sociological' and accounts based on individual opportunity-seeking behavior as 'economic', the paper rejects purely economic accounts on theoretical grounds. Explaining the real workings of institutionalized social norms and social order exclusively in terms of self-regarding opportunity-seeking individual behavior is impossible. An integrated sociological approach to the so-called Hobbesian problem of social order that incorporates opportunity-seeking along with rule-following behavior is necessary. Such an approach emerges on the horizon if economic methods are put to good sociological use on the basis of recent experimental economic findings on rule-following behavior.

Keywords: Hobbesian problem of social order, social institutions, social norms, rational choice theory, rule-following

> "Well, then, says I, what's the use of you learning to do right when it's troublesome to do right and ain't no trouble to do wrong, and the wages is just the same? I was stuck. I couldn't answer that. So I reckoned I wouldn't bother no more about it, but after this always do whichever come handiest at the time." (Huckleberry Finn)

1 Introduction and Overview

Solving the time honored so-called 'Hobbesian problem of social order' (Parsons 1968) requires to account for the existence and maintenance of institutionalized social norms and order *exclusively* in terms of extrinsically motivated individual opportunity-seeking choices of whichever would 'come handiest at the time'.[1] In

1 Opportunity-seeking is used here to mean that all foreseeable causal consequences of a particular act at a particular time and location are taken into account with their likelihoods. Whether an

***Corresponding author: Hartmut Kliemt,** VWL VI, c/o Max Albert, Justus Liebig University, Giessen, Germany, e-mail: hartmut.kliemt@t-online.de

search of an adequate understanding of the emergence and maintenance of institutionalized social norms and order the focus of this essay is on the controversy between *two types of methodological individualists*: those who believe that assuming intrinsically motivated rule-following is indispensable ('sociologists') and those who believe that opportunity-seeking behavior alone can explain the emergence and maintenance of institutionalized social norms and order.[2]

This controversy has been and still is dominated by 'ideal theories' that express what norms 'demand' in 'impersonalized' ways.[3] Yet, "impersonalised statements one might be inclined to make about human societies generally require, if they are to be politically informative, elaboration into statements about particular concrete people doing things to other people" (Geuss 2008, 24).[4] This applies to social norms analogously. They must be upheld' by particular individuals in ways that causally induce particular individuals to show particular behaviors. Only if statistical correlations' or social regularities' in real human behavior are brought about by particular individuals expressing particular predictive and/or prescriptive expectations and complying with them in particular acts can we speak of ('real' or) 'institutionalized' social norms and order.[5]

act has in general certain consequences or what would be the consequences if everybody would perform it does not matter—only the particular act under particular circumstances along with all its foreseeable future causal consequences matters. Intrinsically motivated rule-following behavior takes place if the actor chooses partly independently of the exigencies of a situation according to an envisioned general criterion concerning a class of acts.

2 The paper sidelines the dispute between methodological individualists and adherents of holistic explanations of social phenomena; what I could conceivably say on the topic has been stated in Vanberg 1975.

3 Keeping clear of what has been called 'ideal theory' and 'meta-ethics' is necessary. Feasibility issues are discussed in Brennan/Pettit 2005; Hamlin/Stemplowska 2012; Gaus 2016. 'Institutionalized norms as technology' are discussed in H. Albert 1985 while an excellent older meta-ethical introduction to moral-realism is Sayre-McCord 1988 and a more recent book-length treatment Miller 2013.

4 Jeremy Bentham's observation that 'the demand for a right is no more that right than hunger is bread' (Bentham 1843) elegantly foreshadows Geuss' statement with respect to natural law claims' that attribute existence to such ideal demands without requiring that rights' as real causal factors have to be institutionalized by particular individuals following these demands and showing particular behaviors.

5 The somewhat inflationary use of 'institutional' in this paper is meant to serve as a constant reminder that predictive and prescriptive expectations are addressed 'quid facti' rather than 'quid iuris'. When Vostroknutov speaks of descriptive resp. injunctive norms in this issue of *A&K* his perspective is also strictly quid facti. Within his as well as the moral science perspective of the present paper the corresponding quid iuris questions are not whether in an ideal world ideal norms of a certain type are deemed justified but which of factually viable behavioral technolo-

Most sociologists allow for future-oriented opportunity-seeking behavior along with rule-following. But they typically do not offer a theory of the complementary explanatory roles of opportunity-seeking and rule-following behavior.[6] This is unsurprising, since relying on future-oriented opportunity-seeking as universal behavioral hypothesis prevents economists logically from offering a theory of the complementary role of rule-following. Nevertheless, an understanding of why purely economic accounts of social norms fail can provide insights that are hardly accessible along any other pathway of social inquiry.

The complementary roles that economic and sociological arguments may adopt in integrated accounts of social norms and social order can be illustrated not only by recent game theoretic re-formulations of purely economic approaches but also by their interplay in the history of ideas.[7] Though the history of ideas is not the primary focus of this essay a bird's eye view of the historical role of arguments is suitable to introduce central topics of this essay before I give an overview over the sequence of arguments in the present essay.

It is remarkable that in the recent history of the debate between academic economists and sociologists it went largely unnoticed and is still underappreciated that some sixty years ago the leading legal philosopher of the 20th century, Herbert Hart (see Hart 1961; Hoerster 2013) had rung the death knell on the 'economic theory of law before the (modern) economic theory of law' (see *section 2* below).[8] Yet, in line with the religious tradition that the first knell was rung while the dying person was still alive, there was life in the old dog yet.[9] Since the 1960s advances in economic and game theoretic approaches to institutionalized social norms and order have been breathing new life into the Hobbesian mau-

gies of institutionalized norms seem comparatively more desirable than others according to some factually accepted evaluative standard or other; see for some details and further references Kliemt 2018.

6 The rule-following behavior whose indispensability for an adequate account of social norms sociologists emphasize must, if opportunism is allowed for, be somehow weighed against the ever present 'temptations' to deviate opportunistically from rule following. Human actors can as a matter of fact deviate from envisioned rule-guided prescriptions. So, there needs to be a theory when and how these deviations occur. To the extent that this theory includes the costs of foregoing opportunities the sociologist is already close to using economic *modeling tools* to formulate her own theory.

7 See for an overview of the argument Kliemt 1985 and as a representative anthology Raphael 1969.

8 Hart put together an array of arguments in favor of the central 'sociological' thesis that without intrinsically motivated rule-following the actual workings at least of strictly hierarchical legal institutions could not be adequately accounted for.

9 Ironically Hart rang the knell when so-called economic imperialism took off; see for a popular presentation of the economic imperialism of the time McKenzie/Tullock 1978.

soleum. Formal proofs and explorations of the so-called Folk-Theorem of game theory Robert Aumann (1981) and seminal applications of the underlying ideas by David Lewis (1969), Michael Taylor (1976; 1987) and Andrew Schotter (1981) shifted the 'explanation possibility frontier' of the model of opportunity-seeking inter-active choice making outwards (see *section 3.1*).

Partly overlapping with these approaches which made an extended effort to stay within the constraints of the explanatory model of future-oriented opportunity-seeking choice making, Andrew Schotter's economic theory of social institutions has been a significant step towards incorporating rule-following.[10] Extending David Lewis' original discussion of pure co-ordination (language) games to wider classes of social institutions Schotter was not yet arguing in terms of evolutionary game theory but introduced commitments to *strategies* before evolutionary game modeling made systematic use of such commitments.[11] Evolutionary game modeling created invaluable insights concerning institutionalized social norms and order and managed to popularize them widely.[12] Yet, among economists it went underappreciated that evolutionary models rely on commitments when conceptualizing strategies as *programs* rather than as mere *plans* of opportunistic action (see *section 3.2*).[13]

Ironically it were nevertheless economists like Erik Kimbrough and Alexander Vostroknutov (Kimbrough/Vostroknutov 2016) who presented what seems particularly intriguing and convincing experimental evidence of genuine rule-following among humans and its behavioral effects (see *section 4*). Yet, most economists seem still loath to acknowledge that they are in fact endorsing the central pillar of the behavioral model of individualist sociology if they incorporate rule-following into their models.[14] This is obvious from the fact that they cite Jon Elster's article 'Social Norms and Economic Theory' in the *Journal of Economic Perspectives* as if it were a contribution to economics when Elster states (tongue in cheek): "Ra-

10 This aspect of Schotter's theory is underappreciated in the criticism in Granovetter 1985.

11 See for a more detailed discussion of Schotter's at the time innovative approach Kliemt 1986a and on why Schotter's empirically justified move is problematic within purely economic accounts of institutional in an informal philosophy of law way Kliemt 1987.

12 Impressive examples of this genre are Axelrod 1984; 1986; Voss 1985; Sugden 1986; Schüssler 1985; 1990.

13 In contradiction to the insights of Schelling 1960; Selten 1965.

14 It is another irony of this line of the history of ideas that in particular researchers whose disciplinary affiliations were in economics have been elaborating accounts that were sociological in that they rejected the exclusivity of opportunistic behavior. The most prominent example is, of course, F. A. v. Hayek with V. Vanberg as ally, see e.g. Hayek 1973; 1976; 1979. Formal evolutionary game theory (Maynard-Smith 1982) took off later; Hammerstein/Selten 1994 documents how rapid progress has been.

tionality is essentially conditional and future-oriented. Social norms are either unconditional or, if conditional, are not future-oriented." (Elster 1989, 99)[15]

The two sentences of Elster's statement amount to acknowledging the central difficulty of traditional rational choice approaches to get what is *not* future-oriented, 'social norms', out of 'rationality' which "is essentially [...] future-oriented":[16] In a strictly future-oriented behavioral perspective in which 'bygones are bygones' and rational future-oriented actors 'always do whichever come handiest at the time' there seems no room for unconditional compliance.[17]

Though real human actors do not follow rules whatever the ' (opportunity) costs' of compliance,[18] they do not behave like 'homines oeconomici'. Humans are neither exclusively extrinsically motivated by the *predicted* consequences of responding to the exigencies of each situation of choice-making separately nor are they guided exclusively by envisioned general rules which they interpret as *prescribing* what to do. Acknowledging this, the aim of the subsequent discussion

15 Even an eminent economist like Peyton Young is 'fuzzy' about where his otherwise impressive theories of emergence of social norms transcend the purely economic paradigm; see Young 1998; 2008; 2015.

16 Viktor Vanberg in his excellent work on rules in economics acknowledges that rules cannot be rationally chosen simply because it is desirable to be able to do so (Vanberg 1988, 98ff.). Like Ulysses who needs a mast to be bound to, actors need to be presented with the option of choosing to be committed to a rule. Otherwise they may desire but be unable to choose becoming committed. Insofar there is no substantial disagreement between Vanberg's and the views expressed subsequently. He and I are both endorsing the groundbreaking work on economic methodology of Hans Albert (see H. Albert 1967; 1998; H. Albert et al. 2012). Yet, Vanberg chooses to use the term 'economics' for the type of evidence-oriented discipline Hans Albert envisioned while I think that this tends to perpetuate confusion. This being said I concede that it is, of course, possible to use the term 'economic explanation' in a wider sense that allows for rule-following behavior. In that case the categorical distinction between opportunity-seeking and rule-following has to be made *within* economics. Since I believe that exclusive future-orientation rather than consistency (maximization) is the differentia specifica between economic rational choice and other approaches I chose to draw the line between economic and other accounts of institutionalized social norms in terms of future-orientation.

17 An effort to integrate the strict future-orientation of Austrian economics which would allow for action motivated by the shadow of the future only with the shadow of the past is made in Lewis 2004.

18 Actors are bounded by commitments and not only cognitive constraints in the sense of Simon 1957; 1985. Of course, rules are used by agents to draw conclusions about future events in the light of their knowledge of past events, too. Yet rules as internal prescriptive constraints on choices are different even though they build on the results of predictive uses of rules. Appreciating Wittgenstein, I nevertheless deliberately avoid discussing his concept of rules in any philosophical detail. I will rely instead on the metaphor of a *strategy as program* as opposed to *strategy as plan* allowing for opportunity-cost-dependent deviations from the program; see below.

is to critically assess which central aspects of the economic approach to social institutions should be maintained and which abandoned if the universality claim of the basically Hobbesian economic model cannot be upheld.

The sequence of arguments in the paper will unfold as follows: I start (2.) with characterizing the Hobbesian economic model of opportunity-seeking behavior in relation to ideal-typical alternatives (2.1), illustrate how it has been employed to 'reduce' what appears as rule-following behavior to case-by-case future-oriented opportunity-seeking choice making (2.2) and then hold against it some arguments from Herbert Hart's seminal critique of hierarchical Hobbesian 'economic theories of law' (2.3).[19] In the next section (3.) I will introduce the concept of an equilibrium of mutual threats ('if you beat up my doctoral student then I will beat up your doctoral student!'). I initially focus on the notoriously underappreciated coordinative aspects of social customs in contexts of pure coordination. In these contexts strategic plans will always be executed case-by-case by future-oriented opportunity-seeking actors whose *predictive* expectations are correlated according to the custom (3.1). Introducing mixed, partly coordinative and partly conflictual, motives of opportunity-seeking actors I then explore the limits of intrinsic 'pro-social' distributive motives (whose presence has been demonstrated by numerous game experiments). I conclude that—as indicated already by Hart—beyond pro-social outcome-oriented motives *prescriptive* rules or commitments to the execution of strategic plans for repeated interaction need to be introduced into an adequate account of social norms (3.2). Heeding Bentham's (1843) warning that 'hunger is not bread' and institutions will not be brought into existence by a demand for them I then turn to a crucial experiment by Kimbrough and Vostroknutov (2016; 2018, also Vostroknutov in this issue). The experiment shows that general rule-following dispositions do in fact exist and, in a class of familiar game experiments can indeed account for behavior giving rise to institutionalized norms and regular compliance with them (4.). Final remarks wrap the preceding up with a then hopefully unsurprising outline of how in principle an integrated economic and sociological approach to social norms and social order might conceivably account for their real institutionalized forms in increasingly evidence-based ways (5.).

19 The non-mathematical character of Hart's exemplary critique will hopefully make the arguments easily accessible for those without a background in rational choice theory, RCT, while not boring to death those who are familiar with RCT.

2 Opportunity Taking As If Rule Following Behavior

Though humans conceive of themselves as motivated by rules along with perceived opportunities, a long tradition of social theory tried to explain the workings of social institutions without assuming intrinsically motivated individual rule-following behavior. The aspiration of this strand of theory is to explain regularly occurring overt behavior as resulting from extrinsically motivated opportunity-seeking case by case choice making. To the extent that such efforts succeed, phenomena that commonsense attributes to intrinsically motivated rule-following can be accounted for as resulting from opportunity-taking behavior only.

2.1 Locating the Discussion on the Broader Intellectual Map

With the rise of the modern utility conception since WWII (micro-)economics increasingly allowed for behavior other than that arising from self-regarding extrinsic motives.[20] *Table 1* gives a stylized overview of four possible combinations of 'opportunity-taking—rule-following' and 'selfishness—unselfishness' (indicating exemplary possible 'applications', too):

Tab. 1

Motivational process	**Exclusively opportunity-seeking** [extrinsically or intrinsically motivated uncommitted behavior -- focusing on outcome space]	**Allowing for rule-following** [extrinsically or intrinsically motivated committed behavior -- focusing on strategy space]
Substantive motivation		
Extrinsically-selfish	1 (Standard economics)	2 (Bounded economic rationality)
Intrinsically-unselfish	3 (Benevolent despot politics)	4 (Standard sociology)

20 Classical utility—e.g. hedonistic pleasure and pain—comprises reasons for preferring opportunities, modern utility is not a reason for preferring but represents the ranking of opportunities after reasons have been considered.

The subsequent discussion focuses on what distinguishes the left and the right column.[21] This is not to say, though, that the distinction between the rows is unimportant. In fact, much of the discussion of what is more conventionally understood as 'economic' as opposed to allegedly 'non-economic' motives in social interaction centers around issues of extrinsically-selfish as opposed to intrinsically-unselfish motivation.[22] The subsequent discussion is anchored in cell 1 in which standard economic, extrinsically motivated opportunity-taking behavior is located. 'From there on' it eclectically is extended to cells 2, 3 and 4 as the argument invites and/or requires.[23]

2.2 The Basic Model of Extrinsically Motivated 'Opportunistic Compliance'

Opportunity-seeking can yield a regularity in overt behavior that seems *as if* brought about by intrinsically motivated rule-following provided that in a sequence of separate acts of choice appropriate extrinsic incentives regularly prevail in each and every case. That is, apparently intrinsically motivated rule-following behavior is singled out regularly by extrinsic motives as case-based opportunity-taking choice sequence.[24]

To illustrate concretely, assume that it is deemed desirable that drivers stop at red lights. Assume that a technical device has been implemented in all cars. In each and every instance of passing a flashing red light without stopping, the mechanism sees to it that an electric shock is administered to the driver. Let the shock be sufficiently strong to render passing a flashing red light without stop-

21 For theorists who emphasize the central role of 'commitment power' (Schelling 1960; Selten 1965) the distinction between the columns seems to be the 'nub of the matter'; see on Schelling's role Myerson 2009.

22 See Frey 1997; Bowles 2017 for interesting evidence-based discussions; while Sandel 2012 is a typical example of the philosophical populism and preaching characteristic of some such discussions.

23 Normative ethical theories like certain variants of utilitarianism would be related to behavioral conceptions located in cell 3, other utilitarianisms along with Kantianism in cell 4, while certain variants of virtue ethics would be associated with cell 2 and normative egotism with cell 1; see for some traditional ethical theory background.

24 It may be worthwhile to emphasize that overt behavior which appears as if originating from rule-following behavior cannot 'reveal' that individuals are following a rule unless it is assumed in the first place that individuals *can* follow rules. This assumption would amount to accepting what is classified here as a sociological approach. The intriguing formal exercise of so called case-based decision theory cannot be explored here; see Gilboa/Schmeidler 2003; 2010; 2012 and in an experimental vein Bleichrodt et al. 2016.

ping unattractive for almost all drivers under almost all circumstances.[25] Then, with this 'shocking mechanism' in place, drivers have a sufficiently strong extrinsic motive to stop at flashing red lights. Their behavior will be *as if* guided by a rule or norm to stop at flashing red lights. Once the mechanism is in place, intentional rule-following is not required to explain the observed regularity in overt stopping behavior at red lights. Opportunism in responding to extrinsically motivating incentives—i.e. the expected *regular* electric shock—is all that is necessary to explain the regularity in overt behavior.[26]

Economists and many lawyers have traditionally tried to reconstruct legal institutions as 'mechanisms' that administer extrinsically motivating sanctions along lines analogous to the 'shocking mechanism'. In doing so they treated institutionalized incentive systems as if they could, so to say, be picked from the shelf and then work like pre-programmed machines without further human action. Yet, of course, other than the 'shocking mechanism' of the previous example real institutionalized incentive systems must themselves arise as regularities in human behavior. If *all* behavior is to be explained in terms of opportunistically rational choice making *some human actors* must themselves regularly be extrinsically motivated to administer the sanctions. These actors must then in turn be motivated through sanctions to administer sanctions … and on and on in a hierarchy of sanctions until ultimately a non-sanctioned individual will have to initiate the 'chain reaction'.

2.3 Scope and Limits of Opportunistic Compliance and Hierarchical Norm Enforcement[27]

In the 6th century the Codex Justinianus already raised the question of 'quis custodiet custodes ipsos?' (or in modern parlance: 'who guards the guardians?'). In political theory, the sovereignty conception of absolutism and its idea of the supreme role of an actor who of logical necessity must be 'lege absolutus' (initiat-

25 Even dogs who are controlled by an invisible fence and an electric 'shocking-device' will sometimes 'jump the invisible fence'. Yet, in practically all regular cases the fence will be sufficient to prevent the jump.

26 Provided that drivers are not 'shock lovers' but respond to sufficiently strong electric shocks in the regular manner we expect from our knowledge of 'human nature' the expectation of the shock is sufficient to explain behavior. It may be, though, that there is causal overdetermination in the sense that some drivers are also motivated intrinsically to follow an envisioned rule intentionally; see on the stochastic causality conceptions and the INUS conditions relevant here Mackie 1974; Pearl 2000.

27 The arguments of this section are all inspired by Hart 1961.

ing the chain reaction) eventually grew out of it. This strand of European political theory evolved in step with the evolution of political and legal practice and the growth of the administrative state apparatus.[28] A millennium after the statement in the Codex Justinianus, Bodin's conception of sovereignty emerged.[29] On historically relatively short notice, by mid 17th century Thomas Hobbes extended the underlying logic to the conclusion (Hobbes 1968) that in a pre-institutional 'state of nature'—characterized by the absence of a common 'supreme law-giver'—*every* individual would be 'lege absolutus' (i.e. not-subject to a higher order law-giver).

It is essential here to emphasize that sovereignty is characterized by the *absence of extrinsic motives* operating 'in foro externo'. It does *not* necessarily imply the absence of *intrinsic motives* to forego 'some future apparent good' (Hobbes 1968, chap. 10) if the sovereign individual should choose to do so.[30] That intrinsic motives and commitments to restrict choice making are absent 'in foro interno', too, is an additional premise.[31] Only if this premise is introduced the opportunity-seeking homo oeconomicus of cell 1 of *table 1* becomes the universal explanatory model of individual choice making.[32] If so, individual opportunity-taking action is guided exclusively by the extrinsic motives that arise in pursuit of some 'future apparent good' according to the exigencies of each choice situation taken separately.[33]

---

28 There seem to have been anticipations of this in China much earlier (Fukuyama 2012, part II, chaps. 6–8).

29 Bodin says on the sovereign: "[…] And as the Pope can never bind his owne hands (as the Canonists say;) so neither can a soueraigne prince bind his owne hands, albeit that he would." (Bodin 1992[1576/1606], 92) "If then the soueraigne prince be exempted from the lawes of his predecessors, much lesse should he be bound vnto the lawes and ordinances he maketh himselfe: for a man may well receiue a law from another man, but impossible it is in nature for him to giue a law vnto himselfe, no more than it is to command a mans selfe in a matter depending on his owne will […]." (Bodin 1576/1606, 92)

30 As the first line of central §10, Leviathan states "The 'power of a man,' to take it universally, is his present means, to obtain some future apparent good."

31 The distinction between foro interno and externo is in (Hobbes 1968, chap. 15). Kant teases out some radically Hobbesian implications in his *Metaphysics of Morals* (Kant 1798, §§39–44).

32 Hobbes' 'natural right to everything' is nothing but the absence of any obligation to forego opportunities 'to obtain some future apparent good' unless, of course, self-regarding extrinsic motives suggest otherwise.

33 Taken to its extreme—by adding another premise—extrinsic motivation is the only guidance answering what individuals should and what they would do. What is at stake here is contested to the present day. David Hume had famously claimed that any justification of 'ought' exclusively in terms of 'is', is fallacious. He clearly thought that justifications of what ought to be done are (particular) addressee-relative. They aim at technological advice characterizing the means conducive to particular ends. If certain ends regularly prevail then individuals who are informed about the

In a purely economic account of institutionalized social norms and order the emergence of institutionalized rights and obligations must be explained without recourse to internal commitments of individuals.[34] What this requires with respect to law can be illustrated by John Austin's *The Province of Jurisprudence Determined* (Austin 1954, originally published in 1832).

After subscribing to the separation of law and morals—or to legal positivism in the sense of Hart 1961—Austin focuses on legal institutions.[35] His aim is to separate the law that exists as institutional reality—and as such can exert a causal influence on real behavior (in foro externo)—from the law that is merely desired ('demanded') to exist (in foro interno)—and as such cannot directly exert a causal influence on behavior.[36] This sets Austin's broader agenda: it is necessary, first, to delineate the realm of 'positive' law by discriminating legal from other institutionalized norms according to empirically grounded criteria; second, it needs to be explained exclusively in terms of extrinsically motivated future-oriented opportunity-seeking behavior how institutionalized legal norms can manage to exist as 'positive' law (causally influencing real people).

Austin and many other legal scholars well into the 20th century sought to address both problems simultaneously by explicating the concept of a '(legal) norm' as a command directed by a superior towards an underling. What puts some individuals in the role of superiors (who are *in* command) and puts others in the role of underlings (who are at the receiving end of commands) is in turn explicated in terms of asymmetric power relations.[37] Wishes that are addressed to others become commands if the 'addressor' (who is a potential aggressor) can impose a

means will choose them. In other words what they prudently 'should' and what they will choose tends to coincide—however, *not for logical reasons*. The 'is' and the 'ought' are 'psychologically' intimately related in this tradition.

34 I acknowledge that I am spelling out implications that follow if the Hobbesian logic is taken seriously rather than keeping in line with what is commonly regarded as Hobbes interpretation. But like Spinoza (see Spinoza 1670, chap. 16) who seemed to be inclined towards a radical reading, too, my interests are systematic not exegetic. Hobbes' focus on overt behavior and observable extrinsically motivating incentives is 'behaviorist'. Though Adam Smith was the founder of institutional economics Hobbes was the first 'economic imperialist' in the RCT sense; see Kliemt 2016.

35 "The existence of law is one thing; its merit and demerit another. Whether it be or be not is one enquiry; whether it be or be not conformable to an assumed standard, is a different enquiry." (Austin 1954, 154)

36 Of course, the separation of law and morals is also desirable in view of such ethical and political aims as securing 'nulla poena sine lege' and, more generally, the aim to secure the definiteness/predictability of institutionalized law despite moral pluralism; see for a sophisticated legal philosophy account by a theorist extraordinarily well-versed in economics (Coleman 1985).

37 Hierarchies dominate the politics of all primate societies to some extent (Macchiavelli duly mirrored in Waal 1983). Still, it seems rather far-fetched that ultimately the 'mechanics of power'

sanction negatively affecting the 'utility' expected to be experienced by the 'addressee' *if and only if* the addressee of the wish does not comply with the wish.[38]

That the addressees of commands will comply with the commands if and only if this is in their self-interest in each and every particular case to which the command applies is expressive of the economic behavioral model.[39] But, note, that the behavioral model that applies at the receiving end of commands must also apply to command-giving and enforcing: if the explanatory model of opportunity taking—in its 'cell 1 of *table 1*' incarnation—is assumed to be universal, it must also hold that incurring the costs of executing the threat of sanctions is in each and every case in the self-interest of the individual or institution that administers the sanctions all the way up and down the hierarchical chain of command.[40]

Since the sovereign cannot do all the 'enforcement work' himself he has to rely on 'auxiliaries' forming a 'legal staff'. The auxiliaries must be threatened with sanctions by the sovereign if they do not threaten others with sanctions who then possibly will, in turn, have to threaten still others with sanctions… ad infinitum. Again, the difficulty is that this has to stop with some highest enforcer under the spell of exactly the right extrinsic motives in each and every case of law enforcement. That this be always the case amounts to making a kind of pre-stabilized

---

can be explicated without taking recourse to 'power conferring rules'. This alternative line of attack on a purely economic account of institutionalized social norms and order is closely related to the argument here; see, for an underappreciated Hartian approach to politics in general, the first parts of Barry 1981.

38 The same logic applies to civil rather than penal law issues analogously. "The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it—and nothing else." (Oliver Wendel Holmes 1897 cited after Bowles 2017, 12) In terms of expected utility representations of rational choice making, the 'disutility' of the sanction and the likelihood that it will be *discriminately* administered are assumed to render the expected value of complying with the wishes of the superior a dominant alternative in the choice set of the addressees of the wish. For penal law the expected value model is worked out in an ethical context in Kliemt-Kalweit and Kliemt 1981.

39 From this construction arose what a later particularly perceptive theorist of the relation between extrinsic and intrinsic motivation would call the "duty and interest junction principle: Make it each man's interest to observe […] that conduct which it is his duty to observe." (Bowles 2017, 16). This principle became as *contrary to fact assumption* of 'universal knavery' a central principle guiding policy advice: "That, in contriving any system of government, and fixing the several checks and controuls of the constitution, every man ought to be supposed a knave and to have no other end, in all his actions, than private interest." (Hume 1985, VI/I, 42)

40 In another guise this is nothing but the familiar second-order free-riding problem of public goods theory; see Mueller 2003.

harmony assumption even if we include the possibility that the sovereign has enforcement incentives like a herder who feeds his cows well to get more milk.[41]

Moreover, as far as the legal staff is concerned the role of courts that are assigned the task of *interpreting* prescriptive expectations as semantic entities seems to form another obstacle to an account of law as resulting exclusively from opportunity-taking behavior. In terms of the hierarchical command theory it can conceivably be overcome by providing extrinsic motives to act as interpreter of *rules as semantic entities* in each and every *act of interpretation*: A *sophisticated* sovereign could utter the wish that a certain substantive understanding of a general 'rule-expressive speech act' (an event localized in social space-time) be applied, announce sanctions for non-compliance, and monitor the behavior of judges. From this, an extrinsic motivation of judges to apply the law as expressed through the semantic content of the wish of the sovereign can arise in each and every case in which interpretation is called for.[42] The judge can understand the rule-expressive speech act and use it as an instrument to *predict* the sanctions of the sovereign—or the absence thereof—based on this understanding. *To the extent* that the sovereign can monitor the behavior of judges and can administer sanctions if and only if a judge does not find according to the substantive desire expressed by him as sovereign, judges can fulfill their functions without being intrinsically motivated rule-followers themselves.[43] This considerably extends the scope of the economic account of the workings of an institutionalized legal order.

The preceding is important beyond criticizing '*the economic theory of law before the (modern) economic theory of law*'. It expands our understanding of the ways and means of accomplishing certain ends by a hierarchical social order relying on *predictable* incentives operating as extrinsic motives. With elementary formal tools it can be sketched what kinds of regular behavior—that appear 'as if' caused by rule-following—can be fully accounted for in terms of case-by-case

41 This logic has been explored in 'The power to tax' (Brennan/Buchanan 1980); see also in a 'revisionist' spirit (Brennan/Kliemt 2018). The traditional institutions of 'tax farming' illustrate how much harm is done by implementations of that logic; see also generally Acemoglu/Robinson 2013.

42 A sophisticated sovereign understands what has been aptly called the 'the reason of rules' (Brennan/Buchanan 1985). He cannot himself commit to a rule but intend judges to follow a rule he envisions. He then might in each and every case have an incentive to enforce the content of the rule he intends; see for support Heiner 1983. See for basic criticism of the thesis that this is all there is to 'rule-following' again Kliemt 1987, Vanberg 1988 and the rest of this essay.

43 See this point Baurmann 2009. Of course, the sovereign herself must in each and every case be extrinsically motivated by the exigencies of the situation in which he operates without being subject to a court himself.

opportunity-seeking in a non-hierarchical setting (*section 3.1*). While a *hierarchical* account in which regular compliance is explained exclusively in terms of predicted sanctions must fail on the highest level, mutual 'same-level' sanctioning ('if you beat up my doctoral student then I will beat up your doctoral student!') might still do. According to models of mutual same-level sanctioning, predictable control need not be exerted hierarchically but can conceivably emerge as an equilibrium of *mutual threats* and mutual control of opportunistic behavior. Such a strategic interaction approach does indeed carry a long way towards an economic account of social norms and order, yet, in its pure forms hardly the whole way (*section 3.2*).

3 Predictive and Prescriptive Expectations

Without some rule-following at least on the highest level, the hierarchical threats model of institutionalized legal norms and order can account for the stability of that order only by assuming stability of the extrinsically motivating incentives on the highest level. Only with this natural regularity operative on the highest level, the extrinsically uncommitted enforcer who decides according to the exigencies of each decision taken separately will do so in a predictable way that explains the apparent order.

As mentioned before, the assumption of pre-stabilized natural harmony of interests on the highest level might conceivably be avoided by allowing for mutual threats. To illustrate within a Hobbesian state of nature perspective 'a picture is worth a thousand words':

Fig. 1: With permission from *Munich Social Science Review*, Vol. 1, 28)

The two natives, Crusoe left and Friday right, rely on the threat-potential that their armaments represent vis-à-vis each other.[44] Reciprocal threats of opportunity-seeking actors who are, like Crusoe and Friday, locked into an indefinitely repeated interaction can serve as a functional substitute of a hierarchical sanctioning structure.[45] As long as the sequence of interactions is ongoing they can threaten each other indefinitely. There is no ultimate level of the (time-ordered) 'game-hierarchy' at which the so-called 'shadow of the future' ends.[46] Therefore, opportunity-seeking actors have always to consider 'predictable' future responses of co-actors that their own present choices may trigger.[47]

3.1 Social Customs: Predictive Expectations in Equilibrium

Concretely, imagine that a Crusoe and a Friday of our times—while still castaway—dream of meeting in Berlin exactly a year after being rescued from the island. They have not fixed where to meet in Berlin. When they are—all of a sudden—taken off the island by separate rescue squads they know that they intend to meet in Berlin a year hence. Communication in the meantime is impossible. They know Berlin and know that the other knows the city. Planning on where to go they know that they will meet up at the specified time only if they both plan on going to the same place.[48]

By elementary reasoning about each other's knowledge humans can sometimes as a matter of fact mutually predict their behavior and thereby coordinate

44 *Figure 1* captures the 'natural equilibrium' concept underlying the Buchanan-Bush approach to order in anarchy; see Buchanan 1975 but also MAD, mutually assured destruction, as in Gauthier 1969, append.

45 The literature on reciprocity often relies on fixed computer programs and then studies the evolutionary selection of these programs in a Darwinian spirit. The programs correspond to rule-following rather than future-oriented opportunism. There is a long tradition of such ways of incorporating rule-following into economics; see Alchian 1950, Nelson/Winter 1982, Sugden 1986. In the terminology of this essay these are sociological theories of institutionalized social order and norms.

46 See on some of the formal problems arising from the infinite horizon assumption Güth et al. 1991 and in a philosophy of science perspective Albert/Kliemt 2017.

47 On how in a game model overlapping generations of finitely lived entities can create infinitely lived institutions; see Kandori 1992 and more concretely Brennan/Kliemt 1994.

48 As has been shown in experiments, on Schelling's focal point theory (Schelling 1960) human participants can indeed often manage to coordinate even under information conditions as in the 'meeting in Berlin' example; for related experiments see also Grosskopf/Nagel 2008.

even on a sketchy knowledge basis:[49] For instance, it may be that both speculate that the 'Brandenburger-Tor' will come to mind particularly easily as a prominent meeting point in Berlin. If so, chances are that they will indeed meet up there.

Now, under conditions of ongoing repeated interaction as prevail as long as Crusoe and Friday are still on their island there are much richer sources of mutual prediction available in 'real time' and, for that matter, quite independently of whether or not they can talk to each other. The behavior on former rounds of play will be observable and commonly known to the two actors. Thereby, endogenous to the ongoing interaction, mutually re-enforcing predictions may emerge.

To understand in principle how the process may work out, consider *table 2* below. It illustrates the game form of a 2x2 pure coordination problem (of two players who simultaneously choose one of two moves with no conflict of interest). A has to choose a row and B a column of *table 2* under the assumption that the specific choice does not directly influence the choice of the co-player. Players are assumed to understand the table and the fact that results are co-determined by the causally independent choices of the two choice makers.[50]

If both actors choose, '$C_{(.)}$' the 'substantive payoff combination' (1€, 1€) is the outcome resulting from the 'move combination' or 'profile' (C_A, C_B); if both choose $D_{(.)}$, the move combination (D_A, D_B) with the payoff combination (1€, 1€) is the result of play; if the profile is (C_A, D_B) or (C_A, D_B) the substantive payoff combination resulting as outcome of play will be (0€, 0€).

Tab. 2

A, B choose independently or 'simultaneously', $C_{(.)}$, $D_{(.)}$	C_B	D_B
C_A	(1€, 1€)	(0€, 0€)
D_A	(0€, 0€)	(1€, 1€)

49 Fagin et al. 1995 gives an overview over the logic rather than the empirics of such situations. For related issues of theory absorption as are foundational to game theory, see Leonard 2010 and Güth/Kliemt 2004.

50 For a succinct more technical presentation of essentials of classical game theory particularly relevant to the present concerns, see Albert/Kliemt 2020.

The profiles of moves that lead to corresponding 'substantive payoff profiles' on the main diagonal of *table 2* are in 'equilibrium' if no actor, as long as the move of the other is held fixed can gain by unilaterally moving otherwise. Calling such a situation an 'equilibrium' seems intuitively appealing since none of the opportunity-seeking *individual* actors has a opportunity to make himself better off single-handedly (in terms of the *substantive* outcomes listed as payoff profiles in the table). Each acts in the best way in view of what is already the co-player's best response 'given' his own move.

Using the term 'strategy' for a list of *planned* moves of a player for all contingencies that might arise in a play of a game (see *table 3* below for further illustration of the implications of this definition), we can state that strategic plans are in equilibrium if and only if the plan of each actor is the best response plan to the planned best response of the other for all contingencies that might arise when interacting according to the game form.[51]

Since the strategy concept will be used for an explication of the concept of rule-following a more precise understanding of the strategy concept is necessary.[52] To get an intuitive impression, start by rehearsing the stylized 'pure coordination problem' of *table 2* in which moves and strategies 'coincide'. Actors who are exclusively interested in the substantive monetary result or 'payoff' will rank plays that lead to results on the main diagonal higher than plays that lead to off-diagonal results. If equilibrium profiles—i.e. in the particular case at hand profiles leading to payoff profiles on the main diagonal—prevail the monetary payoffs to each actor are at least as high as those resulting from the non-equilibrium profiles.[53] Therefore, in the special pure coordination case of *table 2*, there is no conflict of interest involved. *Which* equilibrium emerges is irrelevant from the point of view of the two actors. Nevertheless, players face an intricate *planning* problem: They need to form plans leading to one of the equilibria under the constraint that they have to make their ('simultaneous') choices independently of each other without communication (as in the meeting in Berlin case but without

51 In the general case with more than two pure strategies we would have to speak of 'a' best response. Moreover, generalizing to more than two actors would require a few additional concepts without adding any relevant insights to the problems at hand. In equilibrium 'on all levels up' of reflecting on what the actors might do the planned response is the best planned response to the best planned response…

52 On conceptual explication as opposed to factual explanation Carnap 1956; Siegwart 1997.

53 I neglect so-called mixed strategies which do not make much sense here anyway.

the contextual information of that case). They must plan without communication as well.[54]

For solving problems like the preceding 'a change of mind' is insufficient. A change of the game form would help, though. For example, assume, as an instance of a 'sufficient' change of the game form, that—other than in *table 2*—one of the two actors, say A, could move first and B could observe that move before moving herself. Then the game form is as represented in *figure 2*.

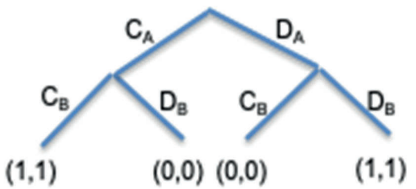

Fig. 2

Note that *figure 2* represents an interaction different from that represented by *table 2* even if the actors' moves are physically the same. According to the rules of interpreting such graphic representations as in *figure 2*, if A has made his move, B, moves knowing A's move. In that case, whatever A would do could and would be 'matched' by B's opportunity-taking choice. For, would A choose move C_A then B would know this and have an incentive to choose C_B. Likewise, D_A would predictively be answered by D_B. Therefore, as long as A expects B to choose as an uncommitted opportunity-taking actor he could predict that *whatever* he, A, would choose, B would as opportunity-seeking actor bring about the co-ordination by her own lights.

To put this slightly differently, as compared with the game form of *table 2* the form presented in *figure 2* bestows a special kind of commitment power on player A. For, according to the rules of the game as represented in the tree of *figure 2*, actor B knows which deed, either C_A or D_A, her co-player has 'committed' in the past. She will be happy to adapt in a way that is advantageous for both in view of the future consequences of her act.

54 Despite the heroic efforts of ideal rational choice theory to explicate 'more geometrico' in general a priori terms—i.e. without referring to empirically 'localized' predictive focal expectations—a standard of forming equilibrium plans for all situations of interactive choice making (Harsanyi/Selten 1988), which of the equilibria will be chosen cannot be answered invoking only a priori rationality principles; see Sugden 1991.

Before A has moved, B cannot move. She can only *plan* how she will move. The possible *plans* of B, are her *four* possible 'strategies' s_{iB}, i=1, 2, 3, 4 which each specify a response planned to be performed after each of A's first moves:

$s_{1B}:=(C_B/C_A, C_B/D_A)$,
$s_{2B}:=(C_B/C_A, D_B/D_A)$,
$s_{3B}:=(D_B/C_A, C_B/D_A)$,
$s_{4B}:=(D_B/C_A, D_B/D_A)$.

As *table 3* shows, some of these *plans* are not reasonable in view of the outcomes. A and B know that B will be able to act opportunistically after A has made a move corresponding to one of his merely two possible plans $s_{1A}:=(C_A)$, $s_{2A}:=(D_A)$. When B actually comes to move, and her strategic plan suggests an alternative to which a better alternative exists, opportunity-seeking future-oriented B will *not* choose according to plan. Whatever A chooses, an opportunity-seeking B will never choose responses that will lead to worse results for her. An A planning to match up with B needs to know only that B is an opportunity-seeking future-oriented choice maker but not the plan of B. Still, what of B's plans?

The next *table 3* represents in the standard interpretation of the so-called strategic form the possible plans and results for the game form:[55]

Tab. 3

	$(C_B/C_A, C_B/D_A)$	$(C_B/C_A, D_B/D_A)$	$(D_B/C_A, C_B/D_A)$	$(D_B/C_A, D_B/D_A)$
C_A	(1€, 1€)	(1€, 1€)	(0€, 0€)	(0€, 0€)
D_A	(0€, 0€)	(1€, 1€)	(0€, 0€)	(1€, 1€)

Inspecting *table 3* shows that—in combination with either $s_{1A}:=(C_A)$ or $s_{2A}:=(D_A)$—the strategy $(C_B/C_A, D_B/D_A)$ will always lead to results at least as good for B as any of her three strategy alternatives.[56] In this sense it is reasonable for A to 'predict'

55 Güth/Kliemt 1995 provides elementary background on terms of strategic as compared to normal form.

56 If this were a game form with prisoner's dilemma like substantive payoffs rather than a pure coordination game the so-called, TFT, Tit-For-Tat, strategy $(C_B/C_A, D_B/D_A)$ would still do well but only in a repeated interaction context. Should B be endowed with the opportunity to commit to a strategy beforehand then it would be in her interest to inform A about this commitment to a

that B will 'plan on matching' whatever he chooses.[57] Yet, as stated already, A need not be afraid that B plans unwisely since nothing can go wrong as long as B is an opportunity-seeking actor who chooses 'whichever come handiest at the time' independently of her plan after learning what A's move is.

In the game form represented by *figure 2* there is an information flow. Obviously, introducing a sequential order among moves of the game form such that former moves are known when later moves are made 'solves' the coordination problem among opportunistically rational actors in such a pure coordination case.[58]

In the real world a transformation of the basic 'one off' game form of *table 2* into the form of *figure 2* may not be viable. Yet, an identical repetition of the basic game form of *table 2* can do the trick.[59] A new game form—a 'supergame form'—emerges from identical repetition. The basic game form of *table 2* itself does not change but it is 'embedded' now into a larger context (becoming one in a sequence of identically repeated forms). Assuming that players have sufficient memory space the behavior on preceding rounds of play can be held 'in memory' by each of the players and 'inform' him on any later round of play. This opens up new causal ways of co-ordination on one of the equilibria on the main diagonal of the basic one-off pure co-ordination problems of the sequence.

To formulate a psychological theory that applies to real behavior in such repeated interactions the reasoning of players would have to be translated into empirical hypotheses about real cognitive processes of real individuals in real time. This empirical study cannot be performed here, yet what has been called 'the logic of the situation'—but in truth would have to be analyzed in realistically complex cases as a 'psychologic'—can be easily analyzed in commonsensical psychological terms as follows: Assume that after a sufficiently large finite number of rounds of play K>>1 a particular sequence of actions has been observed.[60] Since past play has been observed by both players and is 'in memory' this can give rise to a correlation of expectations concerning *future* play as follows: Assume that players

strategy ('rule') and the problem would be solved also in the case of a PD like objective payoff structure; see for details Kliemt 2009, and below.

57 It is highly suspicious that economists tend to refer indiscriminately to all logical implications of their models as 'predictions'. Yet in the case at hand a cognitive psychology explanation plausibly supports this prediction.

58 Due to symmetry in the game form of *table 2* there was no way to coordinate by reasoning alone whereas in that of *figure 2* there was no reason to reflect on coordination at all.

59 Then a way of coordination that replicates certain aspects of the transformation of *table 2* into *figure 2* may emerge from the ability of B to commit to the 'matching strategy' beforehand.

60 That finite memory size may influence viable ways of play and coordination is left out of account here; see for illustration Binmore 1992a and in an elementary constructive way Güth et al. 2005.

chose their moves randomly on each round of interaction until some k, 1<k<K (and without any possibility to communicate directly with each other). According to probability theory individuals will—with sufficiently large k and a probability increasing in k—eventually end up in some 'spell' or sequence of exclusively (C_A, C_B) or (D_A, D_B) plays of the base game form.[61] This subsequence of plays will be yielding (1€, 1€) repeatedly. Let that subsequence of constant play be leading up from k to k'<K; that is k, k+1, k+2, …k'. Without loss of generality, let this subsequence be comprised exclusively of (C_A, C_B).[62] Under plausible assumptions, players will be induced by the observation of the emergent sequence of identical play in the last rounds leading up to k' to conclude that this 'is so for a reason' and *predictively* expect its continuation.[63]

Assume that they expect each other to formulate separately a 'theory' that 'projects' the contingent 'run' of (C_A, C_B) for k+1, k+2, …k' into the future from k'+1… on.[64] This is a crucial empirical hypothesis that links the past to the future despite the fact that for opportunity-seeking actors bygones are bygones. They choose in view of the expected *future* causal consequences contingent on their prediction.[65] That the separate theory formation of separate actors is subject to the same law-like regularities of psychological inductive reasoning (underlying the mental projections of both actors) justifies their individual predictions.[66] Making

61 That base game forms are identically repeated is analogous to that of identical probability distributions in a series of (probabilistically independent) throws of a coin. It is psychologically interesting that human players have comparable difficulties with backward induction (Selten 1978) and constant probability of trials logic.

62 This randomly emergent sequence constitutes a kind of focal point like the Brandenburger-Tor among the castaways.

63 Some roulette players bet on 'ecart'. This is also psychologically caused, yet typically not for a sound empirical reason. The balls do not have a memory that changes, while players' memory changes with history. Unless their memory capacity is zero they cannot 'enter the same stream' (play the same game form) twice.

64 Human actors are biased to see regularities as extending into the future. This is so despite Goodman's new riddle of induction (Goodman 1978). Any function could go on in any way (reminding of Wittgenstein's discussion of rule-following and its non-private character which had to be put aside here altogether; see on the ongoing discussion progressing whether rightly or wrongly so also to coordination games Kripke 1982; Stegmüller 1986; Hacking 1993; Sillari 2013).

65 The game forms of the sequence remain identical like independent roulette throws. If the sequence is infinite even the structure of the series remains identical after the removal of finitely many initial rounds and the adequate equilibrium concept would be subgame consistency then (Güth et al. 1991).

66 If each predicts that the other will apply the same projective theory then each will expect the other to play $C_{(.)}$ on the next round of play after a series of (C_A, C_B). The second order beliefs about the beliefs of the other actor are grounded in empirical theories about the nature of human first

the moves dictated by the empirical 'theories' that 'project' past observations into the future will lead to further play of (C_A, C_B) after k' as a sequence of opportunity taking choices without any commitment.

Without discussing further details, it seems intuitively safe to infer that the individual behavioral projections (models, theories) by which the participants of the interaction form in the shadow of past interactions *predictive expectations* of co-player behavior in future interactions become *causes of non-random regularities in their overt behavior*. The emergent factual regularities are social in the sense that all individual members of a collection of individuals are disposed to form their models interdependently. Yet, they do so separately. The emergent regularity in overt behavior is supported by *mental models* that are common to the participants of the interaction.[67] After an initial phase of adaptation the plans that guide behavior may reach a state of *equilibrium*, in that the interaction "generates messages which do not cause agents to change the theories which they hold or the policies which they pursue" (Hahn 1973, 59).

In sum, in behavioral equilibrium of the pure coordination supergame resulting from identical repetition of the base game form of *table 2* a stable self-supporting regularity of constant play on each round of interaction will emerge. In the terminology adopted for the analytical purposes of this essay a *social custom* has emerged *exclusively on the basis of predictive expectations and opportunity-taking responses to these expectations*. Moreover, not only the emergence but also the maintenance of the specific *social custom* in terms of forward-looking opportunity-seeking choice making can be predicted. But note also that there is no way to form coordinated supergame strategies singling out the particular custom by *planning on a priori grounds* of knowledge of the game form. A contingent particular run of identical plays is necessary to constitute a substantive (focal) expectation on which the formation of inductive predictions can operate to generate as if rule-following behavior on the basis of predictive expectations without any prescriptive expectations.[68]

---

order belief formation and support each other. This is sufficient to provide an extrinsic motive for each actor to plan on $C_{(.)}$ on the next round of play. If the actors as observers of play went on to arbitrary many higher levels in their belief hierarchies they would not come to any other conclusion as long as they share the same theories of human nature and both believe each other to apply the same theories; for related problems see Rubinstein 1989.

67 Yet, it is not only the coincidence of using the same type of models that matters. The participants model aspects of behavior of others involved *in the same interaction* that is known by its common history. Depending on random flux that history could have ended in all C as well as in all D choices.

68 Both the 'run' and the way of 'extrapolating' it, are based on *contingent* facts of individual psychology. They are in the proper sense predictive not 'logical necessities'. The explication of

The insight that we can get thus in descriptive (non-prescriptive) terms, relying on opportunity-seeking behavior and its prediction emerges once *hierarchical* enforcement is substituted by *mutual enforcement* in equilibrium of repeated interactions in pure coordination game forms. This is in itself an extremely important theoretical contribution of supergame conceptions of RCT to our understanding of institutions of social order. Adherents of sociological accounts of social norms and order who dismiss this fundamental insight as a purely formal exercise do so at their own peril. Yet, not all basic game forms are of the pure coordination type (and, even then, in real world evolutionary dynamics will in fact play a role even for predictive expectation formation). The repetition of basic game forms—like in particular the familiar prisoner's dilemma game form—leads to coordination problems among supergame strategies for the whole sequence of repetitions. Even though many assume otherwise the problems of selecting supergame-strategies that coordinate on supergame equilibria are not of the pure coordination type. Going beyond social customs and strategies as plans seems unavoidable.

3.2 Outcome- and Strategy-Orientation

Imagine somebody who plans on travelling abroad. Since he does not have any prior information concerning the ways of driving in the country of destination, he asks beforehand on which side of the road to drive. He receives the information that those who use public roads regularly drive on the left. Treating the information as credible the new prospective new entrant would be induced to share the coordinating predictions of the local drivers. Opportunistic choice making in response to *predictive expectations* that, in the way sketched in the last section, coordinate themselves, seems all that is needed.

Yet, even in cases like choosing the side of the road there may be actors who, say, enjoy the thrill of driving 'against the current' on a busy motorway or who have some particular reason that they would choose to do so at a particular time and place.[69] To cover such outliers even in case of real-world institutions that are in general predictively self-enforcing *prescriptive expectations* to comply with the

social custom given here does not rely on any evolutionary dynamics of strategy selection nor does it invoke normativity of rule following Gibbard 1994.

69 Nozick 1974 provides an early philosophical discussion of the economic theory of law query of why punishment is used as prevention of potentially risky acts even if victims are ex post fully compensated in case the risks actualize themselves. He plausibly makes the argument that those who are not suffering the damages are not compensated for being exposed to the risks.

social custom are typically 'articulated' and sanctioned. When a responder to his query would say to the prospective traveler that the 'rule is to drive on …' her intention typically is to say more than 'the social custom is'. 'I *expect* you to drive on the left' typically expresses something like 'I desire/demand you to drive on the left'[70] and not that 'I predict you to drive on the left and predict that I will myself along with practically all others choose opportunistically to drive on the left, too'.[71]

Within a first-person perspective, the information that people in a foreign country drive on the left is sufficient information for me if I intend merely to avoid collisions when driving there. But I cannot be participating in the local practice of driving in the full sense. To fully participate in that practice it must be possible for me to be guided by 'self-addressing' the prescriptions that typically accompany prevailing social customs (and this is not merely a gentle reminder for those who are imperfectly rational in making predictions about consequences of actions).[72]

Even in what is commonsensically seen as an arbitrary custom a prescriptive element often seems present. To illustrate, recall the iconic scene of Stanley lifting his hat when meeting Livingstone—'out of context' in the middle of Africa—with the words 'Dr. Livingstone, I presume'.[73] The fact that in the illustration of the book on my Grandfather's shelf Dr. Livingstone was depicted responding in kind would have been completely incredible if readers would not beyond the typical hat lifting context have expected intrinsically motivated rule-following guided by (self-addressed) corresponding prescriptive expectations of the two gentlemen participating in the practice prevalent in the community of gentlemen who share not only predictive but also certain prescriptive expectations of what to do.

More importantly, compare the shocking device for controlling drivers with the operation of tax institutions as revenue collecting 'mechanisms'. Though proverbially 'nothing is certain but death and taxes' that tax institutions will apply with as much certainty as the shocking device does not mean that these institutions—shocking as they may be—can be fully understood in predictive

70 As far as the preceding solution of the equilibrium selection problem is concerned 'I predict you to drive on the left' is appropriate.

71 Likewise, if somebody says 'I trust that you will drive on the left' she does not merely express that she '*relies*' on the *prediction* that the co-actor will drive on the left. She alludes to a shared prescriptive expectation. On the notoriously underappreciated fundamental difference between reliance and trust, see Lahno 1995; 2001; 2002.

72 Note in passing that even following a rule of thumb as a guidance is different from case-by-case opportunity seeking adaptation to extrinsic motives as arising from the exigencies of situations taken separately.

73 That in 1968, the band the 'moody blues' published a song 'Dr. Livingstone, I presume' shows how much this scene has become a part of 'folklore'.

terms. The shocking device is '*meant*' to suppress a certain kind of behavior, the taxes are *not meant* to proscribe the taxed behavior. If extrinsic motivation would be all that matters this distinction in meaning could not be made. Even if parking tickets are insignificantly priced and merely sporadically dealt out, we still think that they signal the *prescriptive* expectation not to park in certain places. They are fines not fees. Income-taxes do *not* signal the prescriptive expectation that income earning be omitted.

Predictive expectations play a fundamentally important role in coordinating or 'customizing' real behavior in particular contexts. Yet, when it comes to institutionalized social norms their customary aspect is not all there is. Reaching (self-enforcing) equilibria beyond pure co-ordination problems in general supergame forms requires coordination of supergame *strategies*.

The adoption of these supergame strategies is prescriptively expected independently of whether they are executed in a specific play of the game or not. For instance, the four strategic plans of player B in *table 3* all specify a response for the initial move of A that will not be realized by A's choice. If A would realize C_A, and B plan according to (C_B/C_A, C_B/D_A) then the resulting play of the game would be (C_A, C_B). If B would stick to this strategy after A's initial move D_A the play (C_A, C_B) would be the result, too.

The latter outcome is, as has been argued, excluded if B is opportunity-seeking and (C_B/C_A, C_B/D_A) is merely a plan from which a future-oriented B can deviate after observing D_A. This would change if a prescriptive expectation that B always 'should' perform certain *moves* of, say, a 'C_0-type' could provide (rewards as) extrinsic motives if and only if a specific move of 'C_0-type' has been observed. Yet, prescriptive expectations do in fact seem to range over the full strategies of actors rather than merely moves despite the fact that the provision of extrinsic motives to adopt full strategies seems difficult: Since a full strategy typically specifies responses for unobserved moves it may be directly unobservable which strategy has led to an observed move. For instance, (D_B/C_A, D_B/D_A) and (C_B/C_A, D_B/D_A) may both lead to D_B after D_A while (C_B/C_A, C_B/D_A) and (C_B/C_A, D_B/D_A) may both lead to C_B after C_A. The observation of C_B respectively D_B does not tell which strategy has been driving either.

Obviously intrinsic motivation of an actor could solve the information problem concerning plans. For, excluding self-deception, the actor herself would know her strategic plan including what she would have done according to plan in cases that were not realized. The intrinsic motivation to execute a strategy as a self-addressed prescriptive expectation can conceivably render opportunity-seeking deviation from the plan less attractive than it might have been in case of a strategy without an intrinsic motivation 'attached' to it.

Economists have taken to admitting intrinsic motivation to their models for a rather long time. Yet, giving up extrinsic motivation economists tried to hold up the general outcome-orientation of their models. That economists try to rely on rankings that can be reduced to rankings of outcomes is motivated by their desire to keep the maximization under constraints paradigm intact when. As sketched in the next *section 3.2.1* switching from game forms to games the rankings remain outcome-based. This means that subjective rankings which trace substantive results (classically 'more money' to self is better than less, as in cell 1 of *table 1*) are substituted by subjective rankings that allow for deviations from the 'natural rankings' of substantive outcomes. Many of the impressive results on so-called 'social preferences' of experimental economics are represented as intrinsically motivated preferences over outcomes. Yet, admitting for prescriptive expectations demanding the adoption of complete strategies as commitments—as sketched in *section 3.2.2*—seems necessary.[74]

3.2.1 Game Forms and Games

Initially assume, that actors rank plays of a game exclusively according to the *outcome profiles* that these plays bring about. Such actors are *consequentialists* when forming their rankings of plays of games.[75] Yet, considering *profiles* of monetary payoff consequences in outcome space (rather than merely the self-regarding objective payoff accruing to each evaluator separately) the interpersonal distribution of monetary payoffs can affect rankings. With this information 'envy', 'altruism' etc. can be taken into account in RCT models as represented by subjective rankings. If the rankings of outcomes are compliant with the axioms that guarantee their representability by so-called individual 'utility functions' then the individual utilities that represent rankings of outcome profiles also represent the subjective rankings of plays.[76]

74 In foundational RCT approaches value rankings may range over 'actions' (functions that map states of the world into outcomes) rather than 'outcomes' (e.g. identifying an outcome with a function that maps all states of the world onto that 'constant' outcome); see Gilboa 2009; Savage 1954. But value rankings ranging over strategies are typically avoided.

75 A discussion of important related topics can be found in Broome 1991; 1999.

76 Modern representative utility is a dimensionless ranking rather than measuring a substantive quality like 'pleasure'. The axioms that guarantee that a ranking can be represented by a utility measure trivially assure that the actor behaves *as if* maximizing utility even though there is no intention to maximize utility. The utility index is representing a predictive expectation of what the actor will do. It is—other than the classical utility as motive or reason for ranking and action—

Consider a specific example like the next *table 4*. The tabular representation of the *game form* (left sub-table, in €) is presented side by side with a *game* (right table, in dimensionless 'utilities') that takes into account motives other than self-regarding monetary ones in its subjective payoffs (indicating relative rank of substantive outcome '*distributions*'):

Tab. 4: $\delta = 2/3$[77]. Prisoner's dilemma or pd-game form, 'subjective' game with two equilibria

pd-game form	C_B	D_B		C_B	D_B
C_A	3€, 3€	1€, 4€	C_A	(3, 3) $(u_A, u_B)(3€, 3€) =$ $(f_1(3-\delta \mid 3-3\mid), f_2(3-\delta \mid 3-3\mid))$	(-1, 2) $(u_A, u_B)(1€, 4€) =$ $(f_1(1-\delta \mid 1-4\mid), f_2(4-\delta \mid 4-1\mid))$
D_A	4€, 1€	2€, 2€	D_A	(2, -1) $(u_A, u_B)(4€, 1€) =$ $(f_1(4-\delta \mid 4-1\mid), f_2(1-\delta \mid 1-4\mid))$	(2, 2) $(u_A, u_B)(2€, 2€) =$ $(f_1(2-\delta \mid 2-2\mid), f_2(2-\delta \mid 2-2\mid))$

Start with the left part of *table 4*. It represents a familiar prisoner's dilemma game form. Under the standard 'simplification' of assuming that actors are merely interested in their subjective rankings of the outcome (monetary income) to themselves the unique dominant strategy equilibrium of the *game* is (D_A, D_B) yielding (2€, 2€). Among opportunity-seeking actors whose subjective rankings trace the substantive payoff choosing $D_{()}$ rather than $C_{()}$ is better whatever happens in playing the game. Therefore 2€ is what A as well as B can predict to gain from this interaction separately. In particular, the (C_A, C_B) strategy combination yielding (3€, 3€) is out of their reach as rational opportunity-seeking actors, despite the fact that it would lead to better consequences for each of them.[78]

Assuming that the actors have other-regarding outcome concerns, something like the—merely illustrative particularly simple—subjective rankings of the right

not prescriptively demanding that a certain alternative be ranked higher than another one or 'should' be preferred to another; see for an optimal presentation chapter 2 in Maschler et al. 2013 and classically Herstein/Milnor 1953, additionally Binmore 1992b; Kliemt 2009; Gilboa 2010.

77 As used later δ resp. δ_i correspond with φ, φ_i in the Vostroknutov paper in this issue.

78 These outcomes are out of reach for the same reason that induce A to rely on B choosing the coordinating alternative in the game form of figure 2 whatever the first move of A might be.

part of *table 4* might emerge. The utility functions of the individuals, A, B, in the right part of *table 4* are to be interpreted as representing modifications of natural rankings of results in outcome space.

It is still assumed that only outcomes matter for actors' rankings. Yet, actors do not exclusively focus on monetary outcomes to themselves. They assign some weight to the co-player outcome. The empirical hypotheses concerning the law-like psychological relations that prevail in the ranking processes of the individuals who are confronted with the game form on the left side of *table 4* are represented by the functions f_i, *i=1, 2*. These functions rank results in ways dependent on 'inequality' of outcomes.[79] As shown for illustrative purposes, inequality is measured by the absolute value '|…|' and weighted by $\delta = 2/3$ leading to the rankings depicted in the right part of *table 4*.[80]

Choosing some particularly simple functional forms as in *table 4* and then estimating the parameters is legitimate in principle. Of course, distortions may arise from this as from any simplification. Yet, there may be ways to test the theories underlying f_i, *i=1, 2* and to assess the impact of the distortion on the qualitative validity of model implications.[81]

Adherents of the economic approach who insist on analyzing interactions as games—rather than in terms of game forms—do so because they want to allow for a wider range of motivational factors (including, of course, subjective attitudes to risk). They are aware that motivation need neither be exclusively monetary nor exclusively self-regarding. Moreover, if actors with subjective rankings do in fact make their choices (a) in view of all and only the perceived future causal consequences on *outcome*-profiles of each of the choices taken separately and (b) in line with their subjective rankings of outcomes then choice making can still be 'framed' *as if* individuals were maximizing functions though now their subjective utility functions.

In this framework 'maximization' is *not* as such *an aim* but a consequence of seeking opportunities to realize higher ranked outcomes in each instance of choice making which may be reached in a play of a game this way of presentation has great charms within the rather parochial perspective of (neo-classical)

79 In the example the f_i, *i=1, 2*, 'strip' the dimensionality of '€' away to yield dimensionless individual rankings on the basis of pairs of monetary outcomes and their (in-)equality

80 See on less simplified such approaches which have been foundational for a whole industry of related studies in particular Fehr/Schmidt 1999; Bolton/Ockenfels 2000.

81 It may be noted also that *to the extent* that popular hypotheses concerning so-called 'aversions' [e.g. inequality aversion concerning outcomes] are corroborated by experimental studies or other empirical means the 'utility' representations in the right sub-table of *table 4* are *not* merely ad hoc.

economists: on the one hand, economists can concede that their original assumption of exclusively self-regarding extrinsically motivated seeking of substantive opportunities ('more rather than less money to self') (cell 1, *table 1*) has been refuted for good by the empirical evidence accumulated in experimental economics and experimental psychology, on the other, they can still stick to their basic assumption of opportunity-seeking behavior.

For instance, if in a game experiment with the pd-game form of the left subtable of *table 4* the (C_A, C_B) strategy combination yielding (3€, 3€) is observed economists can admit that the original hypothesis that self-regarding extrinsically motivated opportunity-seeking choice makers would not behave that way is refuted. Yet, they will then turn to the game in subjective payoffs represented in the right part of *table 4*. With the subjective rankings in hand the observed behavior can be presented as aligned with outcome-oriented opportunity-seeking choice making and the standard 'maximization under constraints analytical tools' can still be used with respect to the subjective rankings.[82]

Yet, even though the preceding explains much of the popularity of modifications of utility functions in outcome space *among economists* it does not as such show that this 'behavioral economics' way of adapting economic modeling to empirical evidence is systematically adequate.[83]

In particular, it is not self-evident that it is adequate to focus exclusively on outcome space as in the standard transformations of game forms into games is typically done (the transition from the left to the right of *table 4*). If actors focus on other aspects of the game form than outcomes of play, then it is in no way assured that the rankings of all choices are representable by utility functions defined on outcome space only.

To put it (too) simply, even if in terms of the substantive outcomes of a game form the consequences are the same it matters whether they have been brought about by, say, breaking a promise or keeping it. Whether some person helps another person voluntarily or is compelled to provide help makes a difference for the evaluation of what happens even if the resulting outcome is substantively identical and the person would have acted the same way with or without compulsion.

82 In the case at hand one of the two pure equilibria according to subjective rankings and, for that matter the efficient one, could be selected as result of individual moves.

83 Ad hocery is not the only model risk here. What Hans Albert aptly characterized as Model-Platonism or the proclivity of taking a fictional ideal model for reality itself for no better reasons than sticking to preconceptions of a school and its vested interests is relevant, too (H. Albert 1998; H. Albert et al. 2012). Model-Platonism is endemic not only in economics but also in sociology and in philosophy. In all these fields the proclivity to 'define' the disciplinary agenda on the basis of a priori rather than empirical a posteriori considerations is recurrent.

Though it may be possible to include among the consequences of a choice the nature of the choice itself, this strategy will swiftly become rather complicated. For instance, the fact that the helper has performed an act of voluntary help in the one and an act of coerced help in the other case is with a wide concept of 'consequence' analytically among the consequences of the action taken. Yet, it seems necessary to distinguish between *two relevantly different kinds of consequence* then.[84]

More generally speaking, actors might place value on the play of a game in ways that are non-reducible to rankings of outcome profiles. In particular, they might place value not only on the moves they perform but also on the plans that lead up to these moves.[85] Actors may be intrinsically motivated to place value on conducting certain strategies.[86] To assess the plausibility of such possibilities two meanings of 'strategy' must be distinguished: On the one hand, 'strategy' is used to refer to '*plans* of moves' and, on the other, to '*programs* of moves'. Here 'plan' is assumed to be without any motivational force while the term program is understood as any form of plan that goes along with some motivational force and/or modification of rankings of plays of games (but need not be a program that leaves no choice). This leads back to the core of the controversy between economic and sociological accounts of institutionalized social norms and order.[87]

3.2.2 Strategies as Plans and Strategies as Programs

In classical game theory, a strategy is a *full plan* of how to play a game. For any situation that the planning individual could conceivably be confronted with in

84 For the example of helping, see in a related experimental economic setting Andreoni 1990.

85 To place values on moves or the play resulting from sequences of moves of all players is not discussed here but would be possible in principle as well.

86 As next *section 3.2.2* emphasizes, one should bear in mind that strategic plans specify an intention of how to move even for situations that do not arise in a particular play of a game. For instance, if in the tree of *figure 2*, A chooses option C_A then B's strategic plan (C_B/C_A, D_B/D_A) specifies not only the response C_B but also D_B as response to the alternative move D_A. Even though the branch of the tree starting with D_A is not reached B's strategic plan contains an answer for this like—in larger game trees—all other contingencies (information sets) that can conceivably arise within any particular play.

87 Of course, participants of academic discourse have some leeway to use terms as seems fit to them. Yet, some constraints must be observed. In particular, it is illegitimate to use the same term for two fundamentally different concepts.

the course of playing a game a move is planned.[88] As was rather extensively discussed above, the *strategy as a plan* can be abandoned any time—or 'in action' so to say—by a future-oriented opportunity-seeking player B. Due to this, the concept of *strategy as (complete) plan* is in line with the assumption of opportunistic choice-making. In any instance of choice—technically any information set that can be reached—there is a move planned, yet, the planning individual can at that instance deviate from the plan instantaneously. As opposed to this, the concept of *strategy as program* is incompatible with the assumption of opportunistic choice-making in each and every instance of choice.[89] Like strategies as plans a strategy as program specifies which move will be executed at any instance of choice that may be reached, however, the individual cannot at that instance deviate from the planned move 'at no extra-cost'.

Since the concepts of strategy as plan and strategy as program seem so similar it may be useful to consider again an example from the world of driving to separate the concepts. Imagine that your fancy new autonomous driving car offers you three 'stirring'-options. After you have embarked it in the parking lot: (i) it will move back to your home (with yourself onboard) along a path that you programmed and will do so automatically no matter what; (ii) it will bring you back to your home autonomously along the path that you programmed unless you interfere and stir a course different from the programmed one; (iii) a friendly voice will give you directions corresponding to the program but you must drive—make your choices of where to stir the car—yourself.

If you sit down in the car and choose (i), you are committed to a course.[90] If you choose (ii), you have implemented a kind of 'default' from which you can

88 Recall the difference between moves of B in *figure 2* strategies of B as in *table 3* above. Player B can make two moves, yet, form four different strategic plans that specify the planned response to any first move of A that player B can conceivably learn of in the course of play.

89 What may be called the 'explicitness requirement of non-cooperative game modeling' demands that any aspect of a non-cooperative game model of (inter-)active choice making that is assumed to be beyond the choices of active players must be explicitly modeled as such. In particular intrinsically motivated rule-following and intrinsic other-regarding motives exist for purposes of analyzing a model if and only if explicitly modeled; see on the often overlooked explicitness assumption constitutive for non-cooperative game modeling and its beneficial influence on transparency of economic modeling in more detail Güth/Kliemt 2007.

90 Besides absolute commitments there are also those that simply make alternatives less or more attractive either in subjective or in objective terms; see on this in detail Güth/Kliemt 2007. All commitments require some alterations of what is called as the 'rules of the game'. In the *technical* game theoretic sense the *rules of the game* comprise any aspects of the game that are beyond the influence of choices of the players in any particular play of the game (including the preferences along with the game form). The strategies as plans represent what is *commonsensically* meant by *rules of playing the game*—rules guiding actions and to be followed intentionally.

deviate in response to the exigencies that might arise on the road (psychologically increasing the likelihood that you will follow the course).[91] If you choose (iii), you have self-addressed the prescriptive advice of the friendly voice guiding you more or less like a plan.[92]

According to the model of future-oriented opportunistic choice-making that takes place exclusively in view of the causal consequences of each choice act taken separately, the option (i) differs categorically from options (ii) and (iii). It has introduced the *additional* option of committing to a strategy (and should be explicitly introduced into models as such).[93] This is different from a strategy as mere plan and also—in a graded way—from the options (ii) and (iii). The psychological effect in case of an autopilot of type (ii) will be stronger than that of the friendly voice (iii).[94]

From the point of view of modelling the main message is simple: if *additional options* to commit to strategies as programs as a matter of fact exist they must and can be modeled explicitly as moves in a game tree.[95] Vice versa, once strategies as programs are represented in an extensive form model the necessity to check on their presence in the real world is obvious, too.[96] As far as representing rule-following is concerned the *four* strategies of player B in *table 3* are paradigm ex-

91 Experiments corroborate the behavioral relevance of specifying the default option.

92 It will have a psychological influence on your choices and in this way form a (very weak) commitment, too.

93 In tree-representations of game forms this amounts to adding new branches to the tree; see Kliemt 2009. To link it to another rather popular strand of literature, so-called resolute choice options may or may not exist, yet, whether they exist is a factual issue—hunger is not bread—and should be explicitly modeled McClennen 1990.

94 Even though it should be mere 'noise'—i.e. have no causal effect on case-by-case opportunistic choice-making—the presence of the friendly voice can also make a behavioral difference in the game and then amounts to a change of the game. In terms of standard game theory, the so-called strategic form of the game—in former times often called normal form—invites neglecting the difference between programs and plans. The stenographic device of mapping strategy profiles into outcome profiles (in terms of utilities) conceals that strategies cannot be chosen as moves in the game. They are plans, specifying a move for any contingency that might conceivably arise in a game but are not moves themselves. If players should all play according to plan then, of course, the play of the game is trivially according to plans.

95 A model will be mis-specified unless the options assumed to exist show up explicitly in the game form. It may be worth noting that even if the relevant part of the tree is 'internal' to a personal player it might still be modeled as non-overt behavior by means of decision and game tree models. But quite separate of such apparently fancy possibilities it should be noted that the presence of options may itself be experimentally tested.

96 Engaging a type '(i)' commitment is akin to Ulysses' option to become tied to the mast. Where in case of the metaphor of Ulysses the existence of the mast can be corroborated, in case of other commitments like, say, virtues the factual mechanisms must be checked out; see on the role of

amples of how the content of rules can in principle be modelled.[97] But if strategies are not interpreted merely as plans but as programs of types (i)–(iii) then that different interpretation must be expressed explicitly in the language of RCT, for instance by (additional) branches of a game tree or by modified subjective rankings of its outcomes.

To sum this up, only a concept of a strategy as plan—that does neither modify the ranking of alternatives nor restrict the sets of alternatives that can be chosen—can be fully compatible with the substantive assumption of extrinsically motivated outcome-oriented opportunity-seeking behavior.[98] For the present discussion of extreme (pure) forms of economic accounts of institutionalized social norms it may be noted succinctly: *once strategies are interpreted as programs the rule-following assumption of a sociological approach is implicitly assumed to prevail* (to some extent).[99]

According to the explicitness assumption of non-cooperative game theory only what is *expressed in the language* of RCT is relevant for analytical purposes. As opposed to RCT with its substantive restriction to future-oriented opportunity-seeking and consistent choice making, the underlying *language* of rational choice *modeling*, has room for expressing *both* opportunity-seeking and rule-following behavior. Moreover, with an adequate use of the *language* of RCT the relative importance of opportunity-seeking and rule-following behavior can a. be represented transparently in models and b. studied experimentally if we manage to express commitments to strategies adequately.

---

virtues in the workings of social institutions Baurmann 2002 and on what has been called the strategic role of the emotions which may represent 'unchosen' commitments that do not arise from choices in playing a game see Frank 1988.

97 To model rules as strategies in the strict game theoretic sense is an obvious but somewhat underexplored possibility because assuming that commitments to strategies are possible is alien to non-cooperative game theory which requires that commitments must be represented as additional options in the game form; see on this again in detail (Kliemt 2009). For a related approach that tries to refer to both, strategies pursued from an internal point of view and strategies as descriptions of choices from an external point of view Congleton 2019.

98 Theorists who identify rational choice making with *consistency* of rankings miss this point; see for an example of this the otherwise excellent overview (Diekmann/Voss 2016) where—what I regard as a crucial error—is boldly stated in the final remarks: "Immer wieder hat der Begriff der 'Rationalität' zu Kritik und Missverständnissen geführt. 'Rationalität' ist aber nicht mehr (und nicht weniger) als konsistentes Handeln."

99 In Ken Binmore's terminology the 'eductive' has been given up in favor of an 'evolutionary' approach; see Binmore 1987; any reader interested in further details may consult Kliemt 2009, chap. 5, freely accessible as https://www.uni-giessen.de/fbz/fb02/fb/professuren/vwl/albert/kontakt/mitarbeiter/Kliemt/Buch_1.

Using the *language* of rational choice modeling to bring the empirical (experimental) evidence to bear on understanding institutionalized social norms and order should be the agenda of both sociological and economic approaches. This still unfinished agenda comprises, on the one hand, an exploration of general behavioral facts of the world that individuals 'bring to the table' when they play a particular game and, on the other, how to represent the facts of the particular interaction context adequately. An experiment of Kimbrough and Vostroknutov (Kimbrough/Vostroknutov 2016) on rule-following as such and its application to the institutionalization of social norms may be seen as a major step in pursuit of this agenda (despite the fact that the authors make an effort to present it as if a contribution to conventional RCT).

4 Rule-following and Contextualized Expectations

Kimbrough's and Vostroknutov's experiment on rule-following shows what can be accomplished with experimental methods and ('still') be expressed in terms of apparently conventional economic models formulated in the language of RCT. "The idea is that sociality is driven not directly by preferences over payoff distributions, but rather by preferences for following known social rules […]" (Kimbrough/Vostroknutov 2016, 610). Through the focus on rules the context-dependency that otherwise is hard to overcome by more conventional outcome-oriented social preference accounts of 'prosocial' behavior is mitigated. A substantive prescriptive expectation that is perceived as relevant in a particular context (say equal substantive payoffs, substantive payoffs proportional to contribution, substantive payoffs proportional to need…) is 'plugged' into the *general* proclivity to follow such *contextual* prescriptive expectations.[100]

The perceptions of contextual expectations are captured by Kimbrough and Vostroknutov through a 'norm elicitation' task.[101] To put it very succinctly the sub-

100 My 'only' complaint is that the two authors try to squeeze their insights into the 'utility' framework—going out of their ways to conceal the 'true nature' of their experimental insights by representing them in closed utility functional form. In any event, substantively, they include genuine—intentional—rule following as representing compliance with prescriptive expectations as guiding choices *within* playing a game. The general proclivity to follow rules is itself part of the rules of a contextualized game (in the technical sense).

101 The task is inspired by Krupka/Weber 2013. The use of the term 'norm' here, does not cohere well with the statement: "We model a norm as a strategy profile: a norm describes the most socially appropriate choice for each player in each information set." (Kimbrough/Vostroknutov 2016, 612) The latter use seems more in line with the results of the combination of the general rule

jects are asked what they believe is the prevailing or most common answer to the question of what is prescriptively expected in a situation.[102] To induce participants to think hard, they are informed that they can win a monetary prize if getting close enough to what the other participants in the experiment name as the prevailing view.

Kimbrough and Vostroknutov basically assume that *beliefs* about the substantive content of prescriptive expectations will 'translate' into *actions* according to an individual's *general* proclivity to show rule-following behavior.[103] The 'degree' to which the general proclivity of an individual to follow rules is present is measured by tasks that do *not* involve *inter*-active choice making in *social* situations. In one task actors are, for instance confronted with a screen on which a 'yellow bucket' and a 'blue bucket' are depicted. They are informed that they have 100 'balls' at their disposal which they have to put into either of the buckets. Putting a ball into the yellow bucket will yield 10 ct each—so participants could earn 10.00€ (resp. $) if they would put all balls there—whereas they could earn 5 ct by putting a ball in the blue bin—yielding 5.00€ (resp. $) at max. Actors are informed beforehand: ''the rule is to put the balls into the blue bucket'' (Kimbrough/Vostroknutov 2018, 148). The lower the monetary amount earned, the stronger is the general rule-following proclivity of the participant as measured by the income forgone by following the prescriptive expectation that 'the rule is…' in the rule-following elicitation task.

The guiding modeling idea is that combining the measure of the general rule-following proclivity with an identifier for the perceived particular prescriptive expectations for particular social contexts (as specified by the norm-elicitation task)

following proclivity with the contextualization through specific prescriptive expectations supported here.

102 This is parallel to Crusoe and Friday asking themselves what they know about Berlin and their co-player's knowledge of Berlin when they reflect on 'shared' predictive and prescriptive expectations.

103 This proclivity is assumed to be general in that it is not restricted to particular social contexts (resp. particular types of game forms). In the special context of following 'legal norms' it is for instance very plausible that people who are *general* rule-followers can 'plug in' the particular institutionalized legal norms that contingently prevail in a legal order and then participate in the rule-following practice. Likewise, particular social norms of (positive) moral institutions can be identified and then be followed. The 'other side' of the Hartian 'rule of recognition' which goes beyond empirically identifying the contextually prevailing norms concerns 'actions'. Whether 'I should' comply with a prescriptive expectation recognized to prevail depends on the proclivity to be rule-following. Much more would have to be said here—also with respect to Kelsen's construction of a basic norm in a first-person context… The parallel of the Kimbrough and Vostroknutov approach with well-established philosophy of law constructions seems a corroboration of its potential to unify views on social norms and their institutionalization in law and morals.

leads to improved predictions concerning 'pro-social behavior'.[104] To corroborate that the idea works, Kimbrough and Vostroknutov show that combining the measure of the general proclivity to show rule-following behavior and the elicited beliefs concerning particular prevailing substantive prescriptive expectations can account for pro-social behavior as observed in familiar experimental games.

In the original assault on the problem of explaining social behavior in an earlier paper (Kimbrough/Vostroknutov 2016) the 'proof of concept' is accomplished with a less abstract task for 'measuring' general rule-following inclinations than the 'sorting into buckets task'. In the experiment the (implicit) endowment of participants before entering the first phase of the experiments is €8. After placed in front of a computer terminal participants are confronted with a screen on which five traffic lights are shown (see *figure 3* that is also reprinted in Vostroknutov's paper in this issue as *figure 4*). Participants are informed that for each second that it takes their 'avatar' to pass from starting to end point, €0.08, will be substracted from their initial endowment €8:

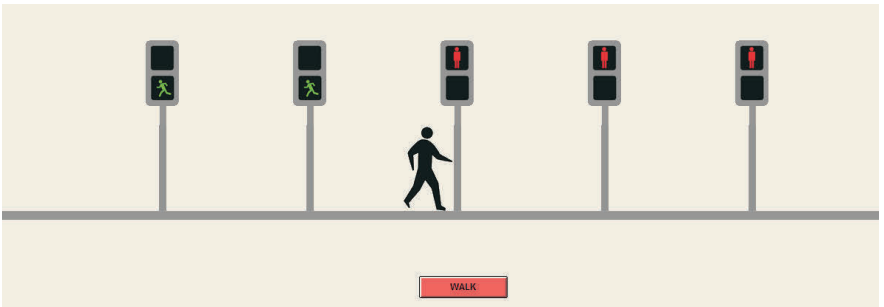

Fig. 3: The rule-following task in Kimbrough/Vostroknutov 2016

Participants learn from the instructions that they can press 'WALK' any time to make the figure on the screen move on. After 'WALK' has been pressed each step takes a second and costs € 0.08. Amounting to €2 in all, the sum of steps will be subtracted from the €8 no matter what. Moreover, participants are informed in the instructions that the figure will stop at each of the lights until it turns green

104 The 'otherworldly' abstractness of the task disentangles the general proclivity to follow rules from substantive prescriptive demands as arise in particular social contexts familiar to the participants. Since the elicitation of prevailing beliefs concerning substantive prescriptive expectations can be meaningful only if the context is familiar the two elicitation processes seem separate.

after another five seconds. If they are not hitting 'WALK'—which would keep the figure moving—this imposes a cost of € 0.08*5=€0.40 on them at each 'Light'.[105]

The information that pressing WALK will keep the figure moving is clear from the instructions. But the instructions state prominently, too: 'The rule is to stop at lights'. The statement of the prescriptive expectation brings about what experimenters call a 'demand effect'. In other experimental contexts exerting such an effect is regarded as a technical blunder. However, for the measuring exercise at hand, the demand effect is a welcome feature of the experimental set up:[106] What is at stake is precisely the potential causal influence of prescriptive expectations as mediated through a general rule-following proclivity.

There are no negative consequences for an actor who is *not* acting in line with the prescriptive expectation contained in the 'demand' that '(t)he rule is to stop at lights'.[107] Quite to the contrary, the actor loses from complying with the demand. Neither are there any negative consequences for other actors to which social preferences in the usual sense could refer.[108] Nevertheless 62,5% of participants waited at least 25 seconds at lights (spending extra time due to reaction time lags) while merely 37,5% were in breach of the rule at least once.

All participants lose at least €2.00 on their way but only those who stop at red lights until they turn green will lose additional amounts of money proportional to the time they wait and the number of red lights at which they stop until those lights turn green. Those who observe all the red lights will lose at least another 2€ (five times €0.40). They end up with at most €4 before entering the next rounds of the experiment. Those who never stop have still €6 of their initial endowment after passing all lights of the first stage of the experiment.

The conjecture is *not* that rule-followers will in all contexts play by the rules, no matter what. Opportunity costs of rule following matter. The conjecture is rather that behavior of those who are more strongly inclined to follow rules than other individuals will generally show significantly more compliant behavior in social interactions in which contextualized prescriptive expectations play a role.

105 "Specifically, we tell subjects to follow a rule, when doing so provides no monetary benefits and instead imposes monetary costs proportional to the time spent following the rule. Under these circumstances, only individuals who are intrinsically motivated to adhere to rules and norms will follow the rule." (Kimbrough/Vostroknutov 2016, 611)

106 See Kimbrough/Vostroknutov 2016, fn. 17, 621.

107 How strong and weak sanctions operate, how self-imposition of demands influences behavior in cases of insufficiently deterrent sanctions etc. is discussed along with other aspects of such settings in Tyran/Feld 2006.

108 "If subjects asked what would happen if they pass through the red light, an experimenter explained that all information relevant to the experiment is in the instructions." (Kimbrough/Vostroknutov 2016, 615, fn. 12)

To illustrate for a stylized experimental setting, assume that there is a 'proposer', *p*, who in the role of a benevolent despot or dictator has to divide a substantive 'pie' of size 2€ between himself, receiving x€, and a recipient, who receives (2-x)€ so that the outcome profile is (x€, (2-x)€). Assume that *p* believes that the prevailing substantive prescriptive expectation relevant in the situation is that of 'equal split'. In the particular case at hand, equal split implies the ideal outcome distribution (1€, 1€) that is x=1€ to self and (2-x)€=1€ to other.[109] This is the game *form* part (corresponding to the left sub-table of *table 4*). To give a description of the corresponding game (corresponding to the right sub-table of *table 4*), Kimbrough and Vostroknutov introduce a subjective ranking or utility function U_p=x(x€)-δ_p*g(|x€-1€|)[110]; where δ_p represents the *general* sensitivity of *p* to rules, x represents subjective *rankings* of substantive opportunities as ranked by self: x(x€)=x;[111] while in δ_p*g(|x€-1€|), the function g(.) is a kind of 'technical scaling' function that operates in combination with δ_p to represent deviations |x€-1€|, from the substantive ideal according to *p*'s subjective rankings.[112]

The ranking of outcomes according to U_p=x(x€)-δ_p*g(|x€-1€|) does not necessarily represent the substantive theories that explain how that ranking arises.[113] Kimbrough and Vostroknutov insinuate otherwise when they discuss U_p=x(x€)-

109 As indicated before, if |x€-1€|=0€ expresses the substantive prescriptive expectation of an ideal distribution relevant in the particular context at hand, then |x-1|€≥0€, 2≥x≥0, measures the substantive deviation from the ideal.

110 Of course, Kimbrough and Vostroknutov treat it as common knowledge that the nature of representative utility is different from classical utility (which treats it as one of the reasons—resp. causes—for ranking) and therefore write more succinctly U_p=x-δ_p*g(|x-1|). This is legitimate to the extent that everybody understands and never forgets that they are talking about functions that *represent* rankings that arise from reasons (motives, causes) that are not necessarily represented term by term in the arguments of the function U_p=U_p(x(x€)-δ_p*g(|x€-1€|))= U_p(x, δ_p, g(.)). The rankings are *approximated* by U_p which is as a matter of scientific practice chosen according to criteria like simplicity, differentiability etc. Whether this function is mis-specified or not is another matter that depends on the underlying law-like regularities and the explanatory aims; see on this from a (critical rationalist) philosophy of science point of view (M. Albert 2013; M. Albert/Kliemt 2017).

111 Note that it would be as well possible to have, say, h(x€)=(x)$^{1/2}$ or some other function that represents how the substantive payoff to self is influencing with some weight—which need not be '1'—the overall ranking of alternative substantive outcome profiles (x€ to self, (2-x)€ to other).

112 Without going into details here it suffices to note that 'x' is a function of substantive payoffs to self—x(x€). It represents the ranking according to self-regarding evaluations of substantive opportunities x€, while δ_p*g(|x€-1€| indicates how detrimental it is from the point of view of *p* if a deviation from the prescriptive demand '1€' of the particular context with a pie of size 2€ arises.

113 The somewhat weird looking x(x€) is used to draw attention to the fact that the function x: X → *R* is needed to map '€' into dimensionless ranking numbers ('utilities') for forming an overall ranking.

δ_p*g(|x€-1€|) determining first order conditions etc. From an *external point* of view of describing overt behavior this is acceptable. Yet, the dictator is *not* making his choice of a particular value of x*€ with the intention of maximizing U_p=x(x€)-δ_p*g(|x€-1€|). That somebody intentionally maximizes a numerical value is psychologically plausible only if that numerical value is expressive of attaining higher levels along some substantive dimension of value (e.g. money or the goods that can be bought with it). Contrary to that the modern utility concept merely represents the position of an alternative and not the substantive dimensions of value that lead to that position in the ranking relative to other alternatives.

Where in classical approaches utility interpreted as, say, hedonistic pleasure (as measured in real quantities of it) is indicating motives for preferring alternatives, modern utility represents only the results of motivational processes.[114] For the motives or other causes that explain how the resulting relative rankings came about we need theories. These substantive theories—classical hedonism being one—rather than the sophisticated curve fitting that leads to U_p=x-δ_p*g(|x-1|) are what matters.

Now, even if it is acknowledged that U_p=x(x€)-δ_p*g(|x€-1€|) is not a theory but merely a representation of rankings whose emergence can be explained by theories it cannot be denied that the components x(x€) respectively δ_p and g(|x€-1€|) correspond to separate theoretical explanatory factors: x(x€) is the influence of substantive payoff x€ to self on rankings, δ_p represents the influence on rankings of a general rule-following proclivity and g(|x€-1€|) that of the localized prescriptive expectation of passing on half of the pie.[115] Thus, even if the particular functional form which adds up strictly self-regarding and other regarding factors (multiplicatively interacting with each other) is not an explanatory theory it still seems that it represents choices in ways that suggest that they are resulting from opportunity-seeking choices that weigh self-regarding and other regarding considerations in each and every instance as represented by the functional form.

It is a matter of mathematical convenience to represent the influences of separate theoretical factors on outcome rankings by a single combined function like

114 To put this slightly otherwise classical, e.g. 'Benthamite', utility theory had it that an alternative was ranked higher *because* it yielded more utility, modern utility theory assigns higher utility because the alternative is ranked higher for what reasons ever; see for the welfare economic implications early on Vickrey 1948.

115 The mapping: 2∼100% 1∼50%, 0∼0% of any substantive pie slightly generalizes the description.

U_p which need not represent termwise the underlying substantive theories.[116] This being said, to separate the general factor of a rule-following proclivity from contextualizing prescriptive expectations as Kimbrough and Vostroknutov do, remains a central achievement.[117]

The dictator game structure mirrors rather well compliance with norms of sanctioning in hierarchical enforcement as discussed in *section 2* of this paper. In *section 3* the mutuality of sanctioning and enforcement was inter-active rather than dictatorial. Kimbrough's and Vostroknutov's discussion is, of course, not confined to hierarchical structures. They do not only discuss experiments in which, as in dictator games, there is only a single active choice-maker but in the main focus of their paper address issues of inter-active, strategic, choice-making. Again, it suffices ti focus on a single paradigm example, so-called voluntary contribution games based on a voluntary contribution mechanism, VCM.[118] (Readers familiar with VCMs should skip the next small print sections.)

For readers who are not acquainted with VCMs the following specific sketch should do as an introduction:

Four individuals i=1, 2, 3, 4 are invited to a laboratory and initially endowed each with $a_1=a_2=a_3=a_4$ =10€. Participants are informed that they must make a choice by which they can contribute any amount b_i=0, 1, …, 10€ to a 'common pool'. Out of this pool,[119] participants i=1, 2, 3, 4 will receive as outcome of play the monetary payoffs $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$: $\pi_i = a_i - b_i + \lambda \sum_{j=1}^{4} b_j$, $0< \lambda <1$, i= 1, 2, 3, 4.

To get an intuitive impression of what is going on set $\lambda = \frac{1}{2}$ and $a_i =: b_i$ for all i= 1, 2, 3, 4. This yields an outcome profile $(\pi_1, \pi_2, \pi_3, \pi_4)$=(20€, 20€, 20€, 20€), then put in 0€ =: b_i, yielding (10€, 10€, 10€, 10€). Obviously, each and every individual is strictly better off if each is contributing everything, b_i=10€, rather than nothing, b_i=0€, to the common project. So each is better

116 That behavior can be represented as if maximizing such a function is not in itself expressive of a causal process (at least no more than describing biological adaptation as if the result of a teleological process expresses that it results from teleological pursuits).

117 This insight projects itself well beyond its original context not only with respect to legal processes. Michael Baurmann drew my attention to relations with the Milgram experiment: Explaining the observed compliance with the demands of the experimenter in the Milgram experiment we can refer to general rule-compliance as contextualized by a first substantive norm of assisting a principal (experimenter) as an agent in executing an experiment and a second contradictory contextual substantive norm of not torturing third parties by ('dictatorially') administering electro shocks. Milgram's observations seem rather well be accounted for as a combined effect of the *general* rule-following proclivity with two 'localized' orthogonal substantive expectations.

118 A VCM is the game form underlying a voluntary contribution game (often also called a public goods game or a tragedy of the commons game; see Hardin 1968; Ostrom 1990).

119 The pool is stocked up with additional amounts of money by the experimenter to model the potential mutual advantages resulting from cooperation by making voluntarily contributions to the common pool.

off if each cooperates. Still, if such interactions are repeated for a number of rounds, typically 10 rounds of play, initially rather high contributions wash out.

At slightly closer inspection potential reasons for this decline of voluntary cooperation/contribution become clearer. Note first, that what each individual will receive as substantive outcome component π_i depends on how much that individual personally contributes as her b_i. Contributing b_i reduces her initial endowment a_i directly to $a_i - b_i$. Of her contribution b_i she will 'get back' $\frac{1}{2} b_i$. This is the fraction λ, $0< \lambda<1$ of the sum of all payments $\lambda \sum_{j=1}^{4} b_j$ which depends on her contribution in that round of play. The rest, $\lambda \sum_{\substack{j=1 \\ j \neq i}}^{4} b_j$, will accrue to her no matter what since it does *not* (causally) depend on her contribution b_i. Therefore, taking each round of the interaction separately it is better for her—and each other participant—not to contribute anything.

According to the rules of the game she can make on each round of play an opportunistic choice of her own b_i that does not directly influence the choices of other individuals. This yields the familiar tension between personal and so-called general interest. However, contributions on an earlier round of play will influence the size of the pie $\lambda \sum_{j=1}^{4} b_j$ that is distributed to the actors. Though actors cannot observe who contributed what, each actor can observe what has been contributed 'in sum'. It is a very robust result of standard experiments of repeated interactions based on VCMs that the sum of contributions monotonically declines if the game is repeated. The best explanation of this is that the prescriptive expectations of participants concerning how much should be voluntarily contributed are not met on earlier rounds of interaction.[120] This may be seen as triggering a kind of 'retributive' response of those who make their own voluntary contributions contingent on meeting their prescriptive expectations.[121]

In a VCM involving 4 participants in which partners are matched with the same co-players for ten rounds of play a decline of contributions from first to last round of play has been robustly observed.[122] Due to their elegant experimental design Kimbrough and Vostroknutov replicate this observation but also reach an interesting contrary result dependent on assortative matching of rule-followers.

Concretely, first 8 individuals are selected from the pool of all subjects. All have participated in the rule-following diagnostic experiment that assigns a particular δ_i, i=1, 2, …, 8 increasing in waiting time to each individual. The 8 players are ordered according to the 'strength' of their δ_i. Then the four higher and the four lower ranking individuals are assigned to two groups to interact for 10 rounds according to a standard VCM (with specific parameters that need not be spelled out here).

The participants in the two groups are not informed about the assortative matching. The game forms are identical for the upper δ and the lower δ groups.

120 There is a kind of end-game effect (Selten/Stöcker 1983). But this does not explain the observed decline.

121 This is the punch line relevant for the present context; see Plott/Smith 2008, sec. 6.1.

122 The decline is observed even if after a so-called 're-start' the actors after a first sequence of interactions are interacting for a second or third such sequence.

Yet, participants of the two groups behave in astonishingly different ways even though they are not informed about the assortative matching.[123] The familiar decline of voluntary contributions (cooperation) does not occur in the groups of high δ rule-followers while it takes place in the groups of those who show lower values of δ in the diagnostic treatment.[124]

Since only 37.5 percent of all participants were not stopping at all lights there must have been quite some individuals with rather high δ in the groups of lesser rule-followers. That the decline took place nevertheless, strongly suggests that the old saying that the first piece of paper on the beach is the worst may apply: those who have a strong proclivity to follow rules (have rather high δ) and are willing to comply with some envisioned substantive prescriptive demand to contribute to VCMs but ended up with other individuals who are not high δ rule-followers tended to reduce their contributions as well.

Leaving out all further detail the following—in view of the existing body of experimental evidence on VCMs quite striking—results are worth emphasizing:

1. For groups of 4 who are rule-followers with lower δ, results are as indicated by the lower curve in *figure 4*. This replicates the conventional results on VCMs in general.
2. For groups of 4 higher δ rule-followers, results are as indicated by the upper curve in *figure 4*. This effect of sorting according to rule-following proclivity refutes general results on VCMs.

The effect of assortative matching deserves special emphasis since this effect has nothing to do with any specific contextualized norm but only with the strength of rule-following per se as matching criterion. This distinct *group level* effect seems out of range of explanations in terms of conventional future-oriented choice making and social preferences of individuals.

123 This matching is done by the experimenter rather than on the basis of signaling etc. as in e.g. Hoppe et al. 2009.

124 The result seems to stand given the data at hand. In view of its surprising—and as I believe—important nature a replication of it—e.g. under restart conditions—seems highly desirable, though.

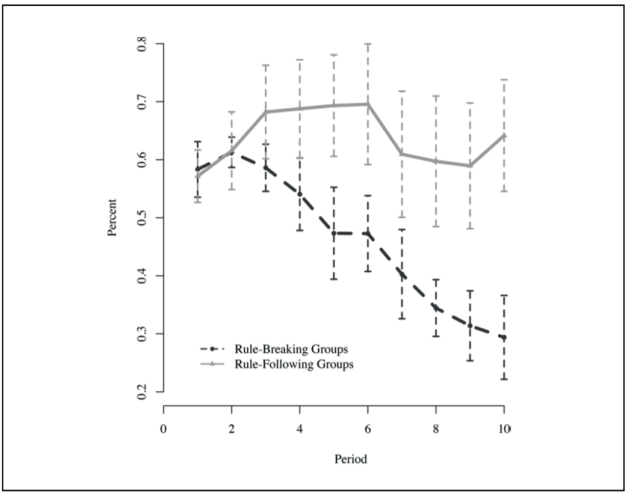

Fig. 4: Kimbrough/Vostroknutov, Norms Make Preferences Social, 623

That under assortative matching deviations from extrinsically motivated opportunism became so much more pronounced than in the previous experimental literature (which relied on random matching across all types) is of great interest. It shows that intrinsically motivated rule-following behavior as such can exert a causal influence on social interaction but also that this influence depends on cofactors like selective matching of rule-following player-types.[125]

In the final remarks I will comment on how these two observations fit into Hume's conception of how social order in large groups arises from small group interactions by which the division of labor is extended in socially embedded processes to the creation and maintenance of institutionalized social order. Before this and some last comments on the climate of discussion among economists and sociologists I will rehearse some central points of the preceding critical assessment of the scope and limits of (purely) economic accounts of institutionalized social norms.

125 A lot more would have to be said here on the contextual nature of the processes involved. It seems significant that homogenous groups of rule followers behave differently in creating substantively favorable outcomes to their members while rule followers in a heterogeneous environment behave more or less like everybody else. It also needs to be further explored whether rule-following proclivities are restricted to certain dimensions of interaction or—like what is indicated by, say, the possibly related but distinct marshmallow experiments—apply across the board to most behavioral dimensions.

5 Economics, Sociology and Social Norms

If we make structural organizational assumptions like that of a 'hierarchy of sanctions', on the one hand, and of 'mutual or reciprocal sanctions', on the other, we can see what can in principle be accomplished if extrinsic motives are administered properly by individuals. Since the organizational structures of sanctioning could not conceivably be mechanical altogether but have to be of the type of institutionalized social orders themselves some kind of rule-following seems indispensable to explain how commitments to execute (positive or negative) sanctions can come into existence.[126] This in turn requires exploring the scope and limits of rule-following.

Herbert Hart put rule-following center stage of the sociology of law but due to the state of experimental work at his time Hart's empirical hypotheses lacked a scientific evidence-based foundation. As far as such a foundation is concerned Kimbrough's and Vostroknutov's paper seems to be a major step in the right direction. The authors manage to 'introduce the experimental method of reasoning' to a study of rule-following per se.[127] 1-They corroborate the thesis that there is a general capacity for rule-following. 2-They show that this capacity does indeed exert a causal influence on behavior in a wide class of games. 3-They show the relevance of the social dimension of assortative matching as a central co-factor in making the causal influence operative in 'collective goods' provision (social order itself being one such good). 4-Finally, their distinction between generalized rule-following (as measured by δ) and the particular prescriptive expectations

126 Mechanisms akin to the shocking device in the case of hierarchical or Leo Szilards doomsday arrangement (removing the remaining opportunistic decisions in MAD) in case of mutual threats have a surreal ring to them but deserve to be explored to understand their limits better. The traditional view of the role of retributive emotions (Mackie 1982) and the so-called altruistic punishment literature, e.g. Fehr/Gächter 2002, are, in a way, substitutes of such mechanisms; see also again in the same spirit Frank 1988.

127 In the times of Hume, the term experimental was not focused as much as today on data gathering methods; see Demeter 2012. That David Hume's *Treatise* was first translated into German language (Hume 1978) by a psychologist is not accidental, though. That Reinhard Selten as one of the founders of experimental economics had already auto-didactically studied theories of empirical psychology before he in the 1950th participated in a game theory course of Ewald Burger (Burger 1966) is also insufficiently known (oral witness Horst Todt) and insufficiently acknowledged. Psychology had a fundamental influence on Selten who always insisted on distinguishing between—in Hobbesian terminology—analyses 'more geometrico' which he compared to theology and diligently pursued with his mathematical skills, on the one hand, and experimental work on the other; see Selten's reponse to Shepsle in Alt et al. 1999 and for some more detail Kliemt 2017.

that (like 'equal outcomes', 'fair contribution') arise from the particular context ('localizing the general proclivity in time and space') seems to capture rather well the 'conventionalist' aspect of institutionalized social norms and order.[128]

Kimbrough and Vostroknutov obviously believe that they can fit their experimental insights into an approach that leaves room for intrinsic motivation but is still compatible with future-oriented opportunity-seeking behavior. For instance, in the dictator game case with a pie of size y, substantive context-dependent prescriptive expectations of $\acute{y}$€ as the appropriate share of the 'pie' (e.g. ½y=$\acute{y}$) are represented in U_p=x(x€)-δ_p*g(|x€- $\acute{y}$€|) in which 'x(x€)' apparently represents the shadow price of norm-following in the ranking function. Thereby it seems that in each instance of decision-making there is an opportunistic choice between complying with the norm or not involved. Yet this impression is deceptive: First, the functional form of representing results is distinct from the theories that explain them. Second, even if U_p=x(x€)-δ_p*g(|x€-$\acute{y}$€|) not only describes overall results but represents the factors leading to these results term-wise, the general proclivity of rule-following (and not only the contextual prescriptive expectation $\acute{y}$€) is part of it. According to the terminology of this paper this renders it a sociological rather than a purely economic approach. Third, this sociological approach is not of the 'oversocialized type' but has a systematic place for particular contextual factors and personal relations (the localizing prescriptive expectations).

As far as an influence of rule-following is concerned economists who acknowledge its presence have argued frequently that it should be treated like friction in physics. Even though frictionless motion does not exist outside the vacuum, friction is negligible and can be left out of account for many purposes and in many specific contexts.[129] Analogously, economists have argued that non-opportunistic behavior occurs only if opportunity costs are very low and would, once it becomes costly, become negligible.[130] More concretely, the factor δ_p*g(|x€-$\acute{y}$€|) is assumed to be very small so that $U_p \sim$x(x€) ('∼' indicating approximately the same). Yet, even if homo oeconomicus behavior and opportunistic future-oriented calculation would always dominate other considerations if costs are 'sufficiently' high this does not show that rule-following at low costs is socially unimportant. Quite to the contrary, human social organization systematically ex-

128 In the coordination game both (C, C) or (D, D) would do. In real life, depending on locality, driving on the left or driving on the right as well as different conceptions of localized justice are relevant; see Elster 1992.

129 For details in a critical rationalist framework; see again M. Albert 2013; M. Albert/Kliemt 2017.

130 In ultimatum experiments actors may leave more than two monthly salaries on the table to sanction offers regarded as too low (Slonim/Roth 1998).

ploits cost-asymmetries:[131] Hierarchy and 'power over' other individuals means basically that some can impose high costs or high benefits on other individuals at low costs to themselves.

In fact Kimbrough's and Vostroknutov's theory underlying the function U_p makes it intelligible why organizational structures put an actor *p* with high δ_p as judge in a situation in which she has low personal stakes in how she finds. An individual with high δ_p is assumed to be rule-following in her interpretation of *prevailing* (contextual) legal prescriptive expectations.[132] Such individuals should be socially sought after if they can be detected.[133]

The effects of assortative matching in the Kimbrough and Vostroknutov experiment are particularly intriguing if we consider selection effects as interacting with cost asymmetries. If we take into account that social order depends on an organizational small group structure this brings the discussion back full circle to Hume. His central thesis concerning social order is that a structure of permanently interacting small groups (the natural organizational form of primates including the human species) is underlying the process of creating and maintaining the institutionalization of social norms and extended social orders of large-scale interaction (Hume 1739, bk. III, Sect. vii).[134]

The trivial but fundamental law of *all* human organization is, that in any ordered large-scale social interaction there is an ordering organizational structure of (internally 'naturally' ordered) small groups of permanently interacting individuals that renders the social order sustainable on ever larger scales.[135] As far as the behavior of the small 'organizing' groups is concerned the combination of individual rule-following proclivities and assortative mixing for collective goods'

131 See on political-organizational effects of low costs Brennan/Lomasky 1985; 1993; Kliemt 1986b.

132 She finds not so much according to what she deems right but to what is the particular law of the land.

133 Pursuing an indirect evolutionary approach allows to systematically address the effects of an ability to discriminate between committed and uncommitted individuals in different contexts; see for specific examples Berninghaus et al. 2003; Brennan et al. 2003; Güth et al. 1999 and for background ideas Frank 1987.

134 It is non-accidental that Hume included this feature of human organization into his account of *human nature*. For him the large scale interaction in the 'company of strangers' (Seabright 2010) that we experience in our modern world is as far removed from the natural adaptation as one can imagine but needs to be explained as resulting from human nature. As mentioned before there are 'ties' to Humean ideas, see also Granovetter 1985.

135 This is an empirical claim concerning the necessity of socially 'embedded' personal small group organizational structures in all large-scale impersonal 'companies of strangers' to which I know no exceptions.

creation in small groups of 'particular' individuals is of interest. By permanent internal personal relations groups of individuals may be committed to courses of collective action in ways individuals would not be.[136] This way the natural proclivities and capacities to follow rules that have evolved in the co-evolutionary process of 'genes, mind and culture' (Lumsden/Wilson 1981) may render increasingly complex social orders feasible and sustainable over extended periods of time.[137] Prescriptive expectations combined with the general proclivity of sufficiently many sufficiently influential individuals to follow rules in ways supported by assortative matching seem key to the astonishing predictability of the workings of social order up to the level of 'Great Societies'.

The preceding remarks on how the division of labor can be extended from small group contexts to the enforcement of rules for larger groups are admittedly vague and speculative. This kind of sociological account of the emergence and maintenance of social order needs further empirical exploration in particular with experimental economic methods.

In fairness to economics as a discipline it must be acknowledged that rule-based deviations from the purely economic account of social norms are implicitly assumed by the broad church of economic theorists who work within an 'evolutionary' framework.[138] Parallel to evolutionary approaches a particularly impressive case in point is the work of Peyton Young which shows how powerful economic theoretical modeling can be.[139] Yet, praising these members of the discipline of economics should not distract of the bad habits that also have been cultivated in the discipline. Becoming increasingly more interested in the prestige and the illusion of universality that mathematical sophistication bestowed on their

136 Rejecting under- and over-socialized accounts of social order and all sorts of utopian anarchism, left or right.

137 The rich literature on the co-evolutionary process relevant here can be found in Henrich 2016.

138 Of an almost endless list of contributions to evolutionary economics Alchian, Hayek, Nelson and Winter, Vanberg have already been mentioned. But there are also economists who have worked on a rather biologically minded evolutionary theory of our species (Bowles/Gintis 2013). A popular presentation of such research which covers also quite some material from experimental economics refers to it as contributions to the 'biology of our species' (Sapolsky 2017). On the more formal side, general modeling tools are presented in Page 2018, in particular chap. 4). On evolutionary game theoretic modeling see Maynard-Smith 1982; Gintis 2000; Hammerstein/Selten 1994; Hofbauer/Sigmund 1984; Schuster/Sigmund 1983; Weibull 1995. More specific techniques particularly relevant in social contexts are simulations of 'game of life' topological structures (Hegselmann 1994; 1996; 2012).

139 An overview of his work cannot be given here since that would require a separate paper. Fortunately, Young himself provided informal accounts of his views on social norms e.g. Young 2008; 2015.

research than in testing the empirical validity of their fundamental hypotheses many economists have been treating empirical arguments and observations of alternative social theories that contradicted their core assumption of opportunity-taking behavior with rather arrogant scorn.[140] To add injury to insult many present day economists seem to claim old insights in particular of social philosophy, social psychology and sociology as *new findings of economics* for no other and better reasons than that they are *new to them*. Willful ignorance of the history of their own field and the traditions and findings of scientific competitors of the economic approach were and still are cultivated in the 'tribe of economists'.[141]

Regrettably much of 'sociology' has also been to a considerable extent about defending certain theoretical preconceptions on a priori grounds (not only by focusing on the issue of holism vs. individualism). In sociology it was not opportunity-seeking, but rule-following behavior guided by internalized norms and values that was taken to extremes of 'over-socialized' conceptions of human behavior (Granovetter 1985). Downplaying opportunity-seeking behavior while focusing more on the discussion of classics of social theory than on theories conducive to empirical research was the result.

The anti-psychological and anti-individualist attitude of many sociologists was no better than the resentment that economists cultivated in opposition to a broadly speaking 'psychological' foundation of their discipline as rooted in laws of human nature.[142] What should have been an ongoing conversation between adherents of economic and sociological approaches to social norms and social order regrettably deteriorated too often into an academic 'shouting contest'. Yet, as recent experimental work shows in the disciples of economics and sociology 'The Times They Are a-Changin'!' (Bob Dylan).

Those interested in more balanced and more extended approaches to social norms and order in a spirit not too far from the one presented here might want to consult (Bicchieri 2006; Bicchieri 2016; Brennan et al. 2013). As far as the role of

140 A brilliant and entertaining sinner's turned victim report is Thaler 2015.

141 In a truly evidence-oriented science in which the results of a long history of trial and error are 'stored' in a canon of corroborated empirical findings concerning theoretical hypotheses (along with empirical evidence and the studies that represent it) the marginal value added by knowledge of alternative development paths that have been abandoned in the history of research may indeed be marginal. However, in economics taking pride in the notoriously short memory of the disciplinary discourse and ignorance of neighboring disciplines is absurd. The more disturbing it is that Daniel Kahneman, a former victim of economists turned Nobelist in economics, in his comments on the topic in Rakow 2010, 463, tries to outperform economists at their worst.

142 On top of this, sociological a priori theories like those of the Frankfurt School cultivated hostility to the fact/value distinction. Large parts of so-called welfare economics do not seem much better, though H. Albert 1958; Sugden 2018.

rules in constitutional political economics is concerned obvious places to look are Vanberg (1994) and Brennan and Buchanan (1985). Following up on 'following the rules' (Heath 2011) and on 'understanding institutions' (Guala 2016) may correct some of my philosophical biases

Acknowledgment: I am indebted to Zombor Meder for the invaluable service of drawing my attention to the experimental work of Erik Kimbrough and Alexander Vostroknutov—presented in Vostroknutov's paper in the same issue of this Journal. The critical comments and inspiration of Max Albert, Michael Baurmann, Anton Leist, Andreas Ryll and Alexander Vostroknutov are gratefully acknowledged. Putting my views into their several economic, sociological and philosophical perspectives helped me to stay the course and to organize my stylized account of economic and sociological approaches to institutionalized social norms around the distinction between opportunity-seeking and rule-following individual behavior.

References

Acemoglu, D./Robinson, J. A. (2013), *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, London

Albert, H. (1958), Das Ende der Wohlfahrtsökonomik, *Gewerkschaftliche Monatshefte 9*

— (1967), *Marktsoziologie and Entscheidungslogik. Zur Kritik der reinen Ökonomik*, Tübingen

— (1985), *Treatise on Critical Reason*, Princeton

— (1998), *Marktsoziologie and Entscheidungslogik. Zur Kritik der reinen Ökonomik*, Tübingen

—/D. Arnold/F. Maier-Rigaud (2012), Model Platonism: Neoclassical Economic Thought in Critical Light, in: *Journal of Institutional Economics* 8, 295–323

Albert, M. (2013), *From Unrealistic Assumptions to Economic Explanations. Robustness Analysis from a Deductivist Point of View* (MAGKS Papers on Economics), Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics

—/H. Kliemt (2017), *Infinite Idealizations and Approximate Explanations in Economics* (MAGKS Papers on Economics), Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics

— (2020), Classical Game Theory, ch. 9.1, in: Knauff, M./W. Spohn (eds.), *The Handbook of Rationality*, Cambridge/MA

Alchian, A. A. (1950), Uncertainty, Evolution, and Economic Theory, in: *Journal of Political Economy* 58, 211–221

Alt, J./M. Levi/E. Ostrom (1999), *Competition and Cooperation: Conversations with Nobelists about Economics and Political Science*, New York

Andreoni, J. (1990), Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, in: *The Economic Journal* 100, 464–477

Aumann, R. J. (1981), Survey of Repeated Games, in: Aumann, R. et al. (eds.), *Essays in Game Theory and Mathematical Economics*, Mannheim, 11–42

Austin, J. (1954), *The Province of Jurisprudence Determined*, London

Axelrod, R. (1984), *The Evolution of Cooperation*, New York
— (1986), An Evolutionary Approach to Norms, in: *American Political Science Review* 80, 1095–1111
Barry, N. (1981), *An Introduction to Modern Political Theory*, London–Basingstoke
Baurmann, M. (2002), *The Market of Virtue*, Dordrecht
— (2009), The Internal Point of View as a Rational Choice?, in: *Rationality, Markets and Morals* 0, 2
Bentham, J. (1843), *Anarchical Fallacies* (J. Bowring, ed.), Edinburgh
Berninghaus, S./W. Güth/H. Kliemt (2003), From Teleology to Evolution. Bridging the Gap Between Rationality and Adaptation in Social Explanation, in: *Journal of Evolutionary Economics* 13, 385–410
Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*, New York
— (2016), *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, Oxford
Binmore, K. (1987), Modeling Rational Players I & II, in: *Economics and Philosophy* 3/4, 179–214/9–55
— (1992a), Evolutionary Stability in Repeated Games Played by Automata, in: *Journal of Economic Theory* 57, 278–305
— (1992b), *Fun and Games—A Text on Game Theory*, Lexington
Bleichrodt, H./M. Filko/A. Kothiyal/P. Wakker (2017), Making Case-based Decision Theory Directly Observable, in: *American Economic Journal: Microeconomics* 9, 123–151
Bodin, J. (1992[1576/1606]), *On Sovereignty*, Cambridge
Bolton, G./A. Ockenfels (2000), ERC: A Theory of Equity, Reciprocity and Competition, in: *American Economic Review* 90, 166–193
Bowles, S. (2017), *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens* (Reprint.), New Haven–London
—/H. Gintis (2013), *A Cooperative Species: Human Reciprocity and Its Evolution* (Reprint.), Princeton
Brandt, R. B. (1959), *Ethical Theory. The Problems of Normative and Critical Ethics*, Englewood Cliffs
Brennan, G./J. M. Buchanan (1980), *The Power to Tax. Analytical Foundations of a Fiscal Constitution*, New York
— (1985), *The Reason of Rules*, Cambridge
—/L. Eriksson/R. E. Goodin/N. Southwood (2013), *Explaining Norms*, Oxford–New York
—/W. Güth/H. Kliemt (2003), Trust in the Shadow of the Courts, in: *Journal of Institutional and Theoretical Economics* 159, 16–36
—/H. Kliemt (1994), Finite Lives and Social Institutions, in: *Kyklos* 47, 551–571
— (2018), Fiscal Powers Revisted: The Leviathan Model After 40 Years, in: Congleton, R. D./B. N. Grofman/S. Voigt (eds.), *The Oxford Handbook of Public Choice*, vol. 2 (English Edition), Oxford–New York–Toronto
—/L. Lomasky (1985), The Impartial Spectator Goes to Washington, in: *Economics and Philosophy* 1, 189–211
—/P. Pettit (2005), The Feasibility Issue, in: Jackson, F./M. Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*, Oxford, 258–279
—/L. E. Lomasky (1993), *Democracy and Decision*, Cambridge
Broome, J. (1991), *Weighing Goods. Equality, Uncertainty and Time*, Oxford
— (1999), *Ethics out of Economics*, Cambridge

Buchanan, J. M. (1975), *The Limits of Liberty*, Chicago
Burger, E. (1966), *Einführung in die Theorie der Spiele*, Berlin
Carnap, R. (1956), *Meaning and Necessity*, Chicago
Coleman, J. L. (1985), *Markets, Morals and The Law*, New York
Congleton, R. (2019), Towards a Rule-Based Model of Human Choice: On the Nature of Homo Constitutionalis, in: *James M. Buchanan: A Theorist of Political Economy and Social Philosophy, ed. R.Wagner* (1st ed. 2018., Bd. J. M. Buchanan), New York, 769–805
Demeter, T. (2012), Hume's Experimental Method, in: *British Journal for the History of Philosophy* 20, 577–599
Diekmann, A./T. Voss (2016), Rational-Choice-Rezeption in der deutschsprachigen Soziologie, in: Moebius, S./A. Ploder (eds.), *Geschichte der Soziologie im deutschsprachigen Raum Handbuch Geschichte der deutschsprachigen Soziologie*, Wiesbaden, 663–682
Elster, J. (1989), Social Norms and Economic Theory, in: *Journal of Economic Perspectives* 3, 99–117
— (1992), *Local Justice. How Institutions Allocate Scarce Goods and Necessary Burdens*, Cambridge
Fagin, R./J. Y. Halpern/Y. Moses/M. Y. Vardi (1995), *Reasoning about Knowledge*, Cambridge/MA–London
Fehr, E./S. Gächter (2002), Altruistic Punishment in Humans, in: *Nature* 415, 137–140
—/K. Schmidt (1999), A Theory of Fairness, Competition, and Cooperation, in: *Quarterly Journal of Economics* 114, 817–868
Frank, R. (1987), If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?, in: *The American Economic Review* 77, 593–604
— (1988), *The Passions within Reason: Prisoner's Dilemmas and the Strategic Role of the Emotions*, New York
Frankena, W. K. (1988), *Ethics* (2nd ed.), London
Frey, B. S. (1997), *Not Just For the Money. An Economic Theory of Personal Motivation*, Cheltenham
Fukuyama, F. (2012), *The Origins of Political Order: From Prehuman Times to the French Revolution*, London
Gaus, G. (2016), *The Tyranny of the Ideal: Justice in a Diverse Society*, Princeton/NJ
Gauthier, D. P. (1969), *The Logic of Leviathan*, Oxford
Geuss, R. (2008), *Philosophy/Real Politics*, Princeton
Gibbard, A. (1994), Meaning and Normativity, in: *Philosophical Issues* 5, 95–115
Gilboa, I. (2009), *Theory of Decision under Uncertainty*, Cambridge–New York
— (2010), *Rational Choice*, Cambridge/MA
Gilboa, I./D. Schmeidler (2003), Inductive Inference: An Axiomatic Approach, in: *Econometrica* 71, 1–26
—/— (2010), *A Theory of Case-Based Decisions*, Cambridge
—/— (2012), *Case-Based Predictions: An Axiomatic Approach To Prediction, Classification And Statistical Learning*, New Jersey
Gintis, H. (2000), *Game Theory Evolving*, Princeton
Goodman, N. (1978), *Fact, Fiction and Forecast* (3. ed.), New York
Granovetter, M. (1985), Economic Action and Social Structure: The Problem of Embeddedness, in: *American Journal of Sociology* 91, 481–510
Grosskopf, B./R. Nagel (2008), The Two-person Beauty Contest, in: *Games and Economic Behavior* 62, 93–99

Guala, F. (2016), *Understanding Institutions: The Science and Philosophy of Living Together*, Princeton

Güth, W./Y. Kareev/H. Kliemt (2005), How to Play Randomly without Random Generator. The Case of Maximin Players, in: *Homo Oeconomicus* 22, 231–255

Güth, W./H. Kliemt (1995), Ist die Normalform die normale Form?, in: *Homo Oeconomicus* 12, 155–183

—/— (2004), Bounded Rationality and Theory Absorption, in: *Homo Oeconomicus* 21, 521–540

—/— (2007), The Rationality of Rational Fools. The Role of Commitments, Persons and Agents in Rational Choice Modeling, in: Peter, F./H. B. Schmid (eds.), *Rationality and Commitment*, Oxford, 124–149

—/B. Peleg (1999), Co-evolution of Preferences and Information in Simple Game of Trust, in: *German Economic Review* 1, 83–110

Güth, W./W. Leininger/G. Stephan (1991), On Supergames and Folk Theorems: A Conceptual Analysis, in: Selten, R. (ed.), *Game Equilibrium Models. Morals, Methods, and Markets*, vol. 2, Berlin, 56–70

Hacking, I. (1993), On Kripke's and Goodman's Uses of 'Grue', in: *Philosophy* 68, 269–295

Hahn, F. (1973), *On the Notion of Equilibrium in Economics*, Cambridge

Hamlin, A./Z. Stemplowska (2012), Theory, Ideal Theory and the Theory of Ideals, in: *Political Studies Review* 10, 48–62

Hammerstein, P./R. Selten (1994), Game Theory and Evolutionary Biology, in: Aumann, R./S. Hart (eds.), *Handbook of Game Theory*, vol. 2, Amsterdam, 929–993

Hardin, G. (1968), The Tragedy of the Commons, in: *Science* 162, 1243–1248

Harsanyi, J. C./R. Selten (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge/MA

Hart, H. L. A. (1961), *The Concept of Law*, Oxford

Hayek, F. A. v. (1973/1976/1979), *Law, Legislation and Liberty: A New Statement of the Liberal Principles of Justice and Political Economy—Rules and Order* (vol.1)/The Mirage of Social Justice (vol. 2)/The Political Order of a Free People (vol. 3), London

Heath, J. (2011), *Following the Rules: Practical Reasoning and Deontic Constraint*, Oxford

Hegselmann, R. (1994), Solidarität in einer egoistischen Welt: Eine Simulation, in: Nida-Rümelin, J. (ed.), *Praktische Rationalität*, Berlin, 349–390

— (1996), Solidarität unter Ungleichen – Eine Computersimulation, in: Hegselmann, R./H.-O. Peitgen (eds.), *Modelle sozialer Dynamiken – Ordnung, Chaos und Komplexität*, Wien, 105–128

— (2012), Thomas C. Schelling and the Computer: Some Notes on Schelling's Essay 'On Letting a Computer Help with the Work', in: *Journal of Artificial Societies and Social Simulation* 15, DOI: 10.18564/jasss.2146

Heiner, R. (1983), The Origin of Predictable Behavior, in: *American Economic Review* 73, 560–595

Henrich, J. (2016), *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter*, Princeton

Herstein, I. N./J. Milnor (1953), An Axiomatic Approach to Measurable Utility, in: *Econometrica* 21, 291–297

Hobbes, T. (1968), *Leviathan*, Harmondsworth

Hoerster, N. (1971), *Utilitaristische Ethik and Verallgemeinerung*, Freiburg–München

— (2013), *Was ist Recht? Grundfragen der Rechtsphilosophie* (2. ed.), München

Hofbauer, J./K. Sigmund (1984), *Evolutionstheorie and dynamische Systeme. Mathematische Aspekte der Selektion*, Berlin–Hamburg

Hoppe, H. C./B. Moldovanu/A. Sela (2009), The Theory of Assortative Matching Based on Costly Signals, in: *The Review of Economic Studies* 76, 253–281
Hume, D. (1739), *A Treatise of Human Nature*, Oxford
— (1978), *Ein Traktat über die menschliche Natur*, Hamburg
— (1985), *Essays. Moral, Political and Literary*, Indianapolis
Kandori, M. (1992), Repeated Games Played by Overlapping Generations of Players, in: *The Review of Economic Studies* 59, 81–92
Kant, I. (1977[1798]), *Die Metaphysik der Sitten*, Frankfurt
Kimbrough, E. O./A. Vostroknutov (2016), Norms Make Preferences Social, in: *European Journal of Political Economy* 14, 608–638
—/— (2018), A Portable Method of Eliciting Respect for Social Norms, in: *Economics Letters* 168, 147–150
Kliemt, H. (1985), *Moralische Institutionen. Empiristische Theorien ihrer Evolution*, Freiburg–München
— (1986a), *Antagonistische Kooperation*, Freiburg–München
— (1986b), The Veil of Insignificance, in: *European Journal of Political Economy* 2/3, 333–344
— (1987), The Reason of Rules and the Rule of Reason, in: *Critica* 19, 43–86
— (2009), *Philosophy and Economics I*, München
— (2016), Economics and Philosophy, in: *Handbook on the History of Economic Analysis: Great Economists Since Petty and Boisguilber*, Cheltenham–Northampton
— (2017), ABC—Austria, Bloomington, Chicago: Political Economy the Ostrom Way, in: Dragos Aligica, P./P. Lewis/V. H. Storr (eds.), *The Austrian and Bloomington Schools of Political Economy*, vol. 22, 15–47
— (2018), On the Nature and Significance of (Ideal) Rational Choice Theory, in: *Analyse & Kritik* 40, 1–29
Kliemt-Kalweit, E./H. Kliemt (1981), Schutz und Gefährdung von Rechten durch die staatliche Kriminalstrafe, in: *Analyse & Kritik* 3, 171–193
Kripke, S. A. (1982), *Wittgenstein on Rules and Private Language*, Cambridge/MA
Krupka, E. L./R. A. Weber (2013), Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?, in: *Journal of the European Economic Association* 11, 495–524
Lahno, B. (1995), Trust, Reputation, and Exit in Exchange Relationships, in: *Journal of Conflict Resolution* 39, 495–510
— (2001), On the Emotional Character of Trust, in: *Ethical Theory and Moral Practice* 4, 171–189
— (2002), *Der Begriff des Vertrauens*, Paderborn
Leonard, R. (2010), *Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900–1960*, Cambridge
Lewis, D. (1969), *Convention*, Cambridge/MA
Lewis, P. (2004), Economics As Social Theory and the New Economic Sociology, in: Lewis, P. (ed.), *Transforming Economics: Perspectives on the Critical Realist*, London, 167–186
Lumsden, C. J./E. O. Wilson (1981), *Genes, Mind, and Culture. The Coevolutionary Process*, Cambridge/MA
Mackie, J. L. (1974), *The Cement of the Universe*, Oxford
— (1982), *Morality and the Retributive Emotions*, in: *Criminal Justice Ethics* 1, 3–10
Maschler, M./E. Solan/S. Zamir (2013), *Game Theory*, Cambridge
Maynard-Smith, J. (1982), *Evolutionary Game Theory*, Cambridge
McClennen, E. F. (1990), *Rationality and Dynamic Choice—Foundational Explorations*, New York–Cambridge

McKenzie, R. B./G. Tullock (1978), *The New World of Economics*, New York
Miller, A. (2013), *Contemporary Metaethics: An Introduction* (2nd ed.), Cambridge
Mueller, D. C. (2003), *Public Choice III*, Cambridge
Myerson, R. B. (2009), Learning from Schelling's Strategy of Conflict, in: *Journal of Economic Literature* 47, 1109–1125
Nelson, R. R./S. G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge/MA
Nozick, R. (1974), *Anarchy, State, and Utopia*, New York
Ostrom, E. (1990), *Governing the Commons*, New York
Page, S. E. (2018), *The Model Thinker: What You Need to Know to Make Data Work for You*, New York
Parsons, T. (1968), Utilitarianism, Sociological Thought, in: Sils, D./R. K. Merton (ed.), *International Encyclopedia of Social Sciences*, New York–London
Pearl, J. (2000), *Causality. Models, Reasoning, and Inference*, Cambridge
Plott, C. R./V. L. Smith (2008), *Handbook of Experimental Economics Results, Vol. 1*, Amsterdam
Rakow, T. (2010), Risk, Uncertainty and Prophet: The Psychological Insights of Frank H. Knight, in: *Judgment and Decision Making* 5, 458–466
Raphael, D.-D. (1969), *British Moralists*, Oxford
Rubinstein, A. (1989), The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge', in: *American Economic Revue* 79, 385–391
Sandel, M. (2012), *What Money Can't Buy: The Moral Limits of Markets*, New York
Sapolsky, R. M. (2017), *Behave: The Biology of Humans at Our Best and Worst*, New York
Savage, L. (1954), *The Foundations of Statistics*, New York
Sayre-McCord, G. (1988), Introduction: The Many Moral Realisms, in: Sayre-McCord, G. (eds.), *Essays on Moral Realism*, Ithaca, 1–23
Schelling, T. C. (1960), *The Strategy of Conflict*, Oxford
Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge
Schüssler, R. (1985), *Struktur und Kooperation* (unpublished MA Thesis)
— (1990), *Kooperation unter Egoisten*, München
Schuster, P./K. Sigmund (1983), Replicator Dynamics, in: *Journal of Theoretical Biology* 100, 535–538
Seabright, P. (2010), *Company of Strangers* (2nd revised ed.), Princeton
Selten, R. (1965), Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit, in: *Zeitschrift für die gesamte Staatswissenschaft* 121, 301–324, 667–689
— (1978), The Chain Store Paradox, in: *Theory and Decision* 9, 127–159
—/R. Stöcker (1983), End Behavior in Finite Prisoner's Dilemma Supergames, in: *Journal of Economic Behavior and Organization* 7, 47–70
Siegwart, G. (1997), Explikation, in: Löffler, W./E. Runggaldier (eds.), *Dialog und System*, Sankt Augustin, 15–45
Sillari, G. (2013), Rule-following as Coordination: A Game-theoretic Approach, in: *Synthese* 190, 871–890
Simon, H. A. (1957), *Models of Man*, New York
— (1985), *Models of Bounded Rationality (1 & 2)*, Cambridge/MA
Slonim, R./A. E. Roth (1998), Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic, in: *Econometrica* 66, 569–596
Spinoza, B. de (1951[1670]), *A Theologico-Political Treatise. A Political Treatise*, New York
Stegmüller, W. (1986), *Kripkes Deutung der Spätphilosophie Wittgensteins*, Stuttgart
Sugden, R. (1986), *The Economics of Rights, Co-operation and Welfare*, Oxford–New York

— (1991), Rational Choice: A Survey of Contributions from Economics and Philosophy, in: *The Economic Journal* 101, 751–785
— (2018), *The Community of Advantage*, Oxford
Taylor, M. (1976), *Anarchy and Cooperation*, London
— (1987), *The Possibility of Cooperation*, Cambridge
Thaler, R. H. (2015), *Misbehaving: The Story of Behavioral Economics*, New York
Tyran, J.-R./L. P. Feld (2006), Achieving Compliance When Legal Sanctions Are Non-deterrent, in: *Scandinavian Journal of Economics* 108, 135–156
Vanberg, V. (1975), *Die zwei Soziologien. Individualismus und Kollektivismus in der Sozialtheorie*, Tübingen
— (1988), Rules and Choice in Economics and Sociology, in: *Jahrbuch für neue politische Ökonomie* 7, 146–167
— (1994), *Rules and Choice in Economics*, London–New York
Vickrey, W. (1948), Measuring Marginal Utility by Reactions to Risk, in: *Econometrica* 13, 319–333
Voss, T. (1985), *Rationale Akteure und soziale Institutionen: Beitrag zu einer endogenen Theorie des sozialen Tauschs* (Reprint 2015), München
Waal, F. de (1983), *Chimpanzee Politics*, London
Weibull, J. W. (1995), *Evolutionary Game Theory*, Cambridge/MA
Young, H. P. (1998), *Individual Strategy and Social Structure. An Evolutionary Theory of Institutions*, Princeton
— (2007), *Social Norms, Department of Economics*, Oxford University, Discussion Paper Series 307
— (2015), The Evolution of Social Norms, in: *Annual Review of Economics* 7, 359–387